



UNIVERSIDAD CARLOS III DE MADRID

TESIS DOCTORAL

Topic Maps y Tecnologías de la Información Jurídica: un modelo de recuperación de información para bibliotecas digitales semánticas jurídicas

Autor:

David García Martul

Director:

José Antonio Moreiro González

DEPARTAMENTO

Biblioteconomía y Documentación

Programa de Doctorado en Documentación

Getafe, abril de 2010

ISBN-10: 84-616-2188-3, ISBN-13: 978-84-616-2188-0

ÍNDICE GENERAL

Índice general	XIII
Índice de figuras	XX
Índice de tablas	XXVII
Índice de gráficas	XXIX
Introducción. Un mapa de conocimiento para la sobrecarga de información.	1
CHAPTER 1. Hypothesis, Objectives and Methodology.	11
1.1 Conceptualization of the research problem.	11
1.2 Hypothesis.	14
1.3 Objectives.	15
1.3.1 Cardinal Objectives.	16
1.3.2 Specific Objectives	18
1.4 Methodology.	19
1.4.1 Contextual Methodological Framework.	19
1.4.2 Qualitative methodology: characteristics and reasons for its use.	24
1.4.3 Qualitative methodology work in our thesis.	25
1.5 Fuentes.	28
1.5.1 Fuentes de información general sobre <i>Topic maps</i> .	33
1.5.2 <i>Topic maps</i> para servicios web.	41
1.5.3 Esquemas, reglas y plantillas.	42
1.5.4 Representación de información con Topic maps.	44

1.5.5 Trabajos sobre la interoperabilidad de <i>Topic maps</i> con metadatos.	44
1.5.6 Diseño de sistemas.	48
1.5.7 Sitios web de información especializada sobre <i>Topic maps</i> .	50
1.5.8 Blogs.	51
1.5.9 Aplicaciones en el mundo real de los <i>Topic maps</i> .	52
1.5.10 Tutoriales.	53
1.5.11 Listas de distribución.	53

BLOQUE I. Modelo Teórico de los *Topic maps* en el contexto de la Arquitectura de Información.

CAPÍTULO 2. Bibliotecas Digitales Semánticas.	55
Introducción.	55
2.1 Concepto y propiedades de las bibliotecas digitales semánticas.	56
2.2 Teoría de la desintermediación.	69
2.2.1 Aspectos que influyen en la desintermediación.	77
2.2.2 La biblioteca como intermediario experto.	80
2.2.3 La teoría de la desintermediación aplicada a la comunidad de usuarios.	82
2.2.4 La teoría de la desintermediación desde el punto de vista de la colección.	82
2.3 Teoría de la organización del conocimiento en bibliotecas digitales semánticas.	84

CAPÍTULO 3. Comunicación de información jurídica en la red.	96
Introducción.	96
3.1 Lenguajes XML en el ámbito jurídico.	97
3.2 Lenguajes de navegación y arquitectura web para sitios jurídicos.	104
3.2.1 Los lenguajes controlados en los sistemas de información jurídica.	106
3.3 El lenguaje de especialidad. El español jurídico.	118
3.3.1 El español jurídico en la informática jurídica documental.	121
3.3.2 Herramientas lingüísticas para el análisis del lenguaje jurídico.	123
3.4 La “Web Semántica” jurídica.	130
3.4.1 Las tecnologías de la información en la “web semántica” jurídica.	134
3.4.2 El lenguaje jurídico en la “web semántica” jurídica.	140
3.4.3 Ontologías y tipos de ontologías jurídicas.	146
 CAPÍTULO 4. <i>Topic maps</i> : Visión Sintética.	 156
4.1 Marco Epistemológico de los <i>Topic maps</i> .	156
4.1.1 Relación entre el Modelo de Información Contextual y los <i>Topic maps</i> .	166
4.1.2 Organización del Conocimiento y Modelo de Información Contextual en la Sociedad Red.	168
4.2 Esquema Básico de Organización de la Información: Índice.	173
4.3 Esquemas para la Organización Semántica de la Información.	176

4.3.1 Taxonomías.	177
4.3.1.1 Metodología en la construcción de taxonomías.	185
4.3.2 Tesauros.	193
4.3.3 Redes semánticas.	204
4.3.4 Ontologías.	206
4.4 <i>Topic maps</i> y Esquemas de Organización Semántica de la Información.	209
4.4.1 Introducción.	209
4.4.2 <i>Topic maps</i> como modelo de organización de contenidos.	216
4.4.3 Concepción holística del modelo <i>Topic map</i> .	232
4.4.3.1 Holística del modelo <i>Topic map</i> en el ámbito jurídico.	242
 CAPÍTULO 5. <i>Topic maps</i> : Visión Analítica.	 256
Introducción.	256
5.1 La navegación contextual.	259
5.2 Origen y normalización de los <i>Topic maps</i> .	263
5.3 Conceptos clave de los <i>Topic maps</i> .	273
5.3.1 Concepto de <i>Topic map</i> .	273
5.3.2 La norma ISO/IEC 13250.	278
5.4 Estructura de los <i>Topic maps</i> .	280
5.4.1 <i>Topics</i> .	284
5.4.2 <i>Occurrences</i> .	287
5.4.3 <i>Associations</i> .	289

5.5 <i>Topic maps</i> en la organización del conocimiento.	292
5.5.1 Los <i>Topic maps</i> en la organización conocimiento.	300
5.5.2 Elementos constitutivos de la organización del conocimiento con <i>Topic maps</i> .	302
5.5.2.1 La relación entre materias y <i>topics</i> .	302
5.5.2.2 La identificación de materias.	303
5.5.2.3 La direccionabilidad de las materias.	304
5.5.2.4 Indicadores e Identificadores de materia.	305
5.5.2.5 Materias Publicadas.	307
5.6 Principios generales para la edición de un <i>Topic map</i> .	312
5.7 Herramientas de edición de <i>Topic maps</i> .	329
5.8 Visualización con <i>Topic maps</i> .	341
5.8.1 Objetivos de la visualización.	345
5.8.2 Modelos generales en visualización de información para <i>Topic maps</i> .	350
5.8.2.1 Grafos y árboles.	350
5.8.2.2 Mapas.	352
5.8.3 Modelos de visualización propios de <i>Topic maps</i> .	354
5.8.3.1 Modelo de visualización en árbol hiperbólico.	355
5.8.3.2 Modelo de visualización en índice.	358
5.8.3.3 Tendencias de visualización con <i>Topic maps</i> .	360
5.9 Aplicación de los <i>Topic maps</i> al ámbito jurídico.	362
5.9.1 Representación esquemática de un <i>Topic map</i> en el ámbito	362

jurídico.

5.9.2 Casos prácticos en la aplicación de *Topic maps* en el ámbito 370

jurídico.

BLOQUE II. Maqueta de aplicación del modelo *Topic map* a una biblioteca digital semántica jurídica. El portal del CEPC.

CAPÍTULO 6. Metodología de construcción de una arquitectura *Topic map* para una biblioteca digital semántica jurídica. 375

Introducción. 375

6.1 Estado de la cuestión sobre metodologías en diseño de *Topic maps*. 376

6.2 El Centro de Estudios Políticos y Constitucionales. 378

6.2.1 Origen y evolución del CEPC. 379

6.2.2 La biblioteca y Centro de Documentación del CEPC a través de su sitio web. 389

6.3 Evaluación global de la biblioteca digital por parte de los usuarios del CEPC. 395

6.3.1 Análisis de resultados de la encuesta. 396

6.4 Propuesta de metodología para la edición de un *Topic map* de Biblioteca Digital Semántica Jurídica. 406

6.4.1 La ontología del *Topic map*. 408

6.4.2 El procedimiento de desarrollo del *Topic map*. 413

6.4.2.1 Personal necesario para la edición y diseño del <i>Topic map</i> .	413
6.4.2.2 El proceso.	415
6.4.2.3 La documentación de la ontología.	420
6.4.3 Guía propuesta para el desarrollo del <i>Topic map</i> .	421
6.4.3.1 Principios generales.	421
6.4.3.2 <i>Topic Types</i> .	421
6.4.3.3 Jerarquía de clases.	423
6.4.3.4 <i>Names</i> .	424
6.4.3.5 <i>Occurrences types</i> internos.	424
6.4.3.6 <i>Occurrence types</i> externos.	425
6.4.3.7 <i>Association Types</i> .	425
6.4.3.8 Identificadores.	427
6.5 Evaluación por parte de los usuarios de la biblioteca digital semántica del CEPC con <i>Topic maps</i> .	428
6.5.1 Marco general de la evaluación final del <i>Topic map</i> .	428
6.5.2 Análisis de resultados de la encuesta.	436
CAPÍTULO 7. Conclusiones.	447
Bibliografía.	451

ÍNDICE DE FIGURAS

Figure 1. Chart of the interrelationship of topic maps in the thesis.	16
Figure 2. Chart of the interrelationship between the scientific fields involved in the development of the thesis.	16
Figura 3. Interrelación de campos informativos empleados.	29
Figura 4. Un ejemplo de Biblioteca Virtual: The WWW Virtual Library.	61
Figura 5. Biblioteca Nacional de Australia. Ejemplo de Biblioteca Híbrida.	63
Figura 6. JeromeDL. Ejemplo de Biblioteca Digital Semántica.	69
Figura 7. Mapa conceptual de búsquedas en la base de datos INSPEC para “Information Analysis”.	86
Figura 8. Mapa conceptual de búsquedas en la base de datos INSPEC para “Content Analysis”.	87
Figura 9. Esquema representativo de la relación de los distintos lenguajes en el marco de la arquitectura de información.	97
Figura 10. Ejemplo de sitio web con búsqueda por etiquetas. El sitio web del CEPC.	116
Figura 11. Resultados de una búsqueda con Wordnet 1.7.1.	124
Figura 12. Ejemplo de una búsqueda con Framenet.	128
Figura 13. Ejemplo de búsqueda por la palabra “casa” en el portal de la administración 060.es.	132
Figura 14. Ejemplo de búsqueda por el tesoro de la administración pública británica: UKAT.	133

Figura 15. Motor de búsqueda del CEPC para el término Historia.	136
Figura 16. Esquema de una arquitectura de información con ontologías.	151
Figura 17. Ideograma del modelo Polyscopic.	161
Figura 18. Esquema representativo de la metodología contextual.	164
Figura 19. Los tres tipos de abstracción en el diseño de información contextual.	165
Figura 20. Gráfico de Pink. Representación temporal de las etapas de desarrollo en función de ATG.	169
Figura 21. Ejemplo de sumario del libro de Jack Park. “XML <i>Topic Maps</i> ”.	174
Figura 22. Ejemplo de taxonomía del Centers for Disease Control and Prevention.	184
Figura 23. Ejemplo de hiperenlace desde el topic “Encefalopatía espongiforme” al sitio web del centro sobre la enfermedad.	185
Figura 24. Ejemplo de 6 formas de acceso a la material MEDLINE en la taxonomía de MeSH.	192
Figura 25. Buscador Visual Thesaurus.	197
Figura 26. Tesauro del Senado. Ejemplo de adaptación del EUROVOC.	200
Figura 27. Ejemplo de representación de relaciones implícitas entre conceptos en el <i>topic map</i> del CEPC.	209
Figura 28. Ejemplo de búsqueda por autor empleando <i>topic maps</i> en el portal NZETC.	214
Figura 29. Representación gráfica de la relación de un <i>topic map</i> con los objetos de información según Garshol.	217

Figura 30. Muestra del <i>topic map</i> para el sitio web del CEPC mostrando el	219
topic correspondiente a la directora del centro.	
Figura 31. Relación de Topic Types para el topic map del CEPC.	220
Figura 32. El <i>Topic instance</i> de la Revista de Administración Pública para	221
el topic map del CEPC.	
Figura 33. <i>Topic Map</i> del CEPC donde se recoge el topic type del	225
catálogo de revistas.	
Figura 34. Arquitectura de un <i>topic map</i> engine: Omnigator.	226
Figura 35. Ejemplo de búsqueda de información tributaria con el Tax Map.	234
Figura 36. Ejemplo de interfaz gráfico drop-down list para un topic map.	237
Figura 37. Diseño de la arquitectura de una ontología jurídica: Proyecto	248
Lois.	
Figura 38. iFAQ. Interfaz de búsqueda para IURISERVICE.	249
Figura 39. Portal organizado con <i>topic maps</i> del gobierno noruego.	250
Figura 40. Glosario de Derecho Comunitario del Portal Europa.	252
Figura 41. <i>Topic Map</i> para el Centro de Estudios Políticos y	255
Constitucionales.	
Figura 42. Esquema del modelo matemático de información de Shannon.	261
Figura 43. Cuadro resumen de los estadios de desarrollo para la	269
aprobación de normas ISO.	
Figura 44. Esquema general de la familia de normas <i>Topic map</i> .	271
Figura 45. Representación esquemática de un <i>Topic map</i> .	274
Figura 46. Esquema de las secciones constitutivas de la norma ISO/IEC	280
13250.	

Figura 47. Representación de los <i>topics</i> en un <i>Topic map</i> para un sistema gestor de conocimiento.	284
Figura 48. Representación de las occurrences como flechas en un <i>Topic map</i> .	287
Figura 49. Las <i>associations</i> en el contexto de un <i>Topic map</i> .	289
Figura 50. Representación de una <i>association</i> entre dos <i>topics</i> .	289
Figura 51. Esquema del <i>Topic map</i> como organizador del conocimiento.	304
Figura 52. Ejemplo de tesauro editado con Omnigator por Pepper.	320
Figura 53. Mapa conceptual editado con la herramienta TheBrain para el diseño de la ontología <i>Topic map</i> del CEPC.	322
Figura 54. Ejemplo de diseño de un schema <i>Topic map</i> con el programa ONOTOA.	323
Figura 55. Muestra de diseño del <i>topic map</i> para el Centro de Estudios Políticos y Constitucionales.	333
Figura 56. Ejemplo de <i>topic map</i> de reyes de Inglaterra creado con Hypergraph.	336
Figura 57. Diagrama generado con Thinkgraph.	337
Figura 58. Ejemplo de <i>topic map</i> con la aplicación <i>Topic map designer</i> .	338
Figura 59. Ejemplo de formulario para la edición del <i>basename</i> de un topic con TMtab.	340
Figura 60. Presentación de los resultados de búsqueda con el mapa conceptual Kartoo.	344
Figura 61. Presentación de los resultados de búsqueda con el <i>topic map</i>	345

para Kodak.

Figura 62. Representación visual de la información para el sitio web del CEPC con el editor de mapas conceptuales Touchgraph.	352
Figura 63. Representación en árbol.	357
Figura 64. Página principal del <i>topic map</i> sobre Ópera.	360
Figura 65. Representación que muestra el modelo propuesto.	362
Figura 66. Representación de los <i>topic maps</i> en relación con la colección.	363
Figura 67. Representación gráfica de un <i>topic map</i> sobre LPAC.	368
Figura 68. Ejemplo de <i>topic map</i> para publicaciones sobre derecho tributario.	371
Figura 69. Arquitectura de edición del <i>topic map</i> Tax Map.	372
Figura 70. Ejemplo del sitio web de acceso al <i>topic map</i> para el ciudadano desde Internet.	374
Figura 71. Boletín Oficial del Estado donde se publica la constitución del Instituto de Estudios Políticos.	381
Figura 72. Publicación de los fines del instituto.	382
Figura 73. Organización del instituto.	382
Figura 74. Disposición final por la que se ordena el depósito de los fondos de las Cortes en la biblioteca del nuevo instituto.	383
Figura 75. Orden por la que se crea la Escuela de Administración Pública.	384
Figura 76. Real Decreto sobre organización y funcionamiento del Centro de Estudios Constitucionales.	386
Figura 77. Real Decreto sobre organización y funcionamiento del Centro	387

de Estudios Políticos y Constitucionales.

Figura 78. Página web del listado de revistas publicadas por el CEPC en la pestaña “Publicaciones”.	390
Figura 79. Aplicación RAP empleada para la búsqueda de un artículo.	391
Figura 80. Resultado de la búsqueda del artículo.	392
Figura 81. Recuperación del artículo a texto completo.	392
Figura 82. Centro de Recursos del CEPC.	393
Figura 83. Página web con información sobre el Diploma de Estudios Políticos y Constitucionales con su correspondiente mapa de sitio en el margen izquierdo de la página.	394
Figura 84. Encuesta digital realizada entre los usuarios.	396
Figura 85. Clasificación de <i>topic types</i> en Tax Map.	409
Figura 86. Clasificación de <i>topic types</i> en el <i>topic map</i> del CEPC.	409
Figura 87. Mapa conceptual para el diseño de la ontología <i>topic map</i> del CEPC.	411
Figura 88. Visualización del <i>topic map</i> del CEPC con Vizigator.	411
Figura 89. <i>Topic map</i> de tesauros diseñado por Pepper.	412
Figura 90. Esquema de las fases de desarrollo de la ontología <i>topic map</i> según la metodología de esta tesis.	415
Figura 91. Ejemplo de notas recogidas en el mapa conceptual.	417
Figura 92. Ejemplo de topic que no es <i>topic type</i> .	422
Figura 93. Ejemplo de <i>topic type</i> .	422
Figura 94. Ejemplo de PSI para un <i>topic type</i> .	427

Figura 95. Estructura en dos niveles de acceso a los artículos a texto completo de la Revista de Estudios Políticos editada por el CEPC.	430
Figura 96. Cuestionario final que se envió a los usuarios con mayores habilidades.	433
Figura 97. Interfaz principal de la biblioteca digital semántica para el CEPC.	434
Figura 98. Pantalla resumen de elementos del <i>topic map</i> para el CEPC.	435

ÍNDICE DE TABLAS

Table 1. Diferencias entre metodología cualitativa y cuantitativa.	22
Tabla 2. Actitud del bibliotecario ante la desintermediación	72
Tabla 3. Grupos de trabajo del Consorcio Oasis para la aplicación del lenguaje XML al ámbito jurídico.	101
Tabla 4. Principales proyectos de ontologías jurídicas desarrolladas hasta la fecha.	153
Tabla 5. Método <i>top-down</i> .	188
Tabla 6. Método <i>bottom-up</i> .	190
Tabla 7. Tabla de las principales ontologías jurídicas publicada por Valente.	244
Tabla 8. <i>Topics</i> .	365
Tabla 9. <i>Associations</i> .	366
Tabla 10. <i>Roles</i> .	367
Tabla 11. Tramo de edad de los usuarios encuestados.	397
Tabla 12. Nivel de estudios de los usuarios encuestados.	398
Tabla 13. Lenguas de comprensión lectora para de los usuarios.	399
Tabla 14. Dominios de interés para los usuarios encuestados.	400
Tabla 15. Periodicidad de acceso a la biblioteca del CEPC.	401
Tabla 16. Periodicidad de acceso a la biblioteca digital del CEPC.	402
Tabla 17. Respuesta de los usuarios sobre facilidad de uso de la biblioteca.	403
Tabla 18. Respuesta de los usuarios sobre el grado de dificultad de la búsqueda en la biblioteca digital del CEPC que se les pidió realizar.	404
Tabla 19. Respuesta de los usuarios sobre el grado de dificultad de la	405

búsqueda que se les pidió realizar en el catálogo.

Tabla 20. Tarjeta de evaluación cualitativa de los recursos del CEPC.	417
Tabla 21. Resultados de la pregunta acerca de la impresión global de la tecnología <i>Topic map</i> .	436
Tabla 22. Respuestas de los usuarios a la localización de revistas del CEPC con <i>Topic maps</i> .	437
Tabla 23. Respuestas de los usuarios sobre si los <i>topic types</i> utilizados son adecuados para estructurar la biblioteca digital semántica del CEPC.	438
Tabla 24. Respuestas de los usuarios respecto a la idoneidad de los <i>topics</i> empleados.	439
Tabla 25. Respuestas de los usuarios a la idoneidad de las <i>occurrences</i> empleadas para los <i>topics</i> seleccionados.	440
Tabla 26. Respuestas de los usuarios respecto de la idoneidad de las <i>associations</i> empleadas.	441
Tabla 27. Respuestas de los usuarios a la comparación entre el <i>topic map</i> del CEPC y el <i>topic map</i> del NZETC.	442
Tabla 28. Respuestas de los usuarios a la comparativa entre el <i>topic map</i> del CEPC y el IRS <i>Tax Map</i> .	444
Tabla 29. Respuestas de los usuarios sobre visualización de los tres casos de <i>topic maps</i> mostrados.	445
Tabla 30. Respuestas de los usuarios a la posibilidad de demandar la implantación de <i>topic maps</i> en la biblioteca digital jurídica del CEPC.	445

ÍNDICE DE GRÁFICAS

Gráfica 1. Gráfica correspondiente a los tramos de edad de la comunidad de usuarios.	397
Gráfica 2. Gráfica correspondiente a los niveles educativos de la comunidad de usuarios.	398
Gráfica 3. Gráfica correspondiente a la comprensión lectora de los usuarios en distintos idiomas.	399
Gráfica 4. Gráfica correspondiente a las áreas a las que se adscriben los usuarios del CEPC de acuerdo con sus demandas informativas.	400
Gráfica 5. Gráfica correspondiente a la frecuencia con la que los usuarios acuden físicamente a la biblioteca del CEPC.	402
Gráfica 6. Correspondiente a la frecuencia de acceso a la biblioteca digital del CEPC. Elaboración propia.	403
Gráfica 7. Correspondiente a la facilidad que encuentran los usuarios para el acceso a la biblioteca digital del CEPC.	404
Gráfica 8. Correspondiente al grado de dificultad de las búsquedas en la biblioteca digital del CEPC planteadas al usuario.	405
Gráfica 9. Correspondiente al grado de dificultad de las búsquedas en el catálogo planteadas al usuario.	406
Gráfica 10. Sobre la impresión global de la tecnología <i>Topic map</i> .	437
Gráfica 11. Sobre la localización de revistas del CEPC con <i>Topic maps</i> .	438
Gráfica 12. Sobre si los <i>topic types</i> utilizados son adecuados para estructurar la biblioteca digital semántica del CEPC.	439

Gráfica 13. Sobre la idoneidad de los <i>topics</i> empleados.	440
Gráfica 14. Acerca de la idoneidad de las <i>occurrences</i> empleadas para los <i>topics</i> seleccionados.	441
Gráfica 15. Sobre la idoneidad de las <i>associations</i> empleadas según los usuarios.	442
Gráfica 16. Acerca de la comparación realizada por los usuarios entre el <i>Topic map</i> del CEPC y el <i>topic map</i> del NZETC.	443
Gráfica 17. Sobre respuestas de los usuarios a la comparativa entre el <i>Topic map</i> del CEPC y el <i>IRS Tax Map</i> .	444
Gráfica 18. Sobre preferencias de visualización entre los tres casos de <i>Topic maps</i> mostrados a los usuarios.	445
Gráfica 19. Acerca de la posibilidad de demandar la implantación de <i>Topic maps</i> en la biblioteca digital jurídica del CEPC.	446

Introducción. Un mapa de conocimiento para la sobrecarga de información.

La organización y gestión de la información para apoyar los procesos de investigación y aprendizaje de nuestra comunidad de usuarios es la misión primordial del profesional de la Documentación¹.

Las profundas transformaciones que la tecnología y el trabajo con la red está provocando en esos procesos, así como un nuevo concepto de unidad de información de naturaleza compleja, que está surgiendo modelada en las nuevas prácticas de investigación y aprendizaje, están cuestionando la capacidad de los sistemas tradicionales de gestión de la información en los centros de documentación para tratar esa complejidad y soportar los procesos de trabajo de sus usuarios en el nuevo entorno digital basado en la web y las posibilidades de colaboración que ofrece².

Durante la mayor parte de su historia, los sistemas de información han vivido en el mundo de lo impreso. Todo nuestro sistema de gestión se basaba en un mundo en el que la información era estática y lineal. Estaba pensada para ser utilizada por las personas en lugares concretos. El sistema era gestionado por eruditos capaces de la descripción de unos objetos con portadas, contraportadas, sumarios, índices y colofones. Con el tiempo las técnicas de descripción se han ido puliendo hasta la constitución de lo que hoy conocemos como las técnicas de catalogación, integradas en el corpus conceptual de la Ciencia de la Información³.

La primera gran revolución fue la automatización del catálogo que generó la primera información digital que entró en las bibliotecas: los registros MARC o metadatos. Y esto supuso un cambio que quizás en su momento pasó desapercibido y que hoy se muestra

¹ LÓPEZ YEPES, J. El nuevo profesional de la información, del conocimiento y de la comunicación: El bibliotecario universitario. *Anales de Documentación*, nº10, 2007, p.263-279.

² Especialmente a través de lo que se ha dado en llamar la Web 2.0 o Web Social, en donde cualquier usuario puede participar en la edición de contenidos a través de blogs, o bien en la definición de un vocabulario controlado compartido como son las folksonomies.

³ ESCOLAR SOBRINO, H. *Historia de las bibliotecas*. Madrid : Fundación G. Sánchez Ruipérez, 1990.

como clave para la reconceptualización del sistema⁴. Teníamos una información que no estaba pensada para ser leída únicamente por las personas, sino fundamentalmente por las máquinas. A partir de ahí y con la penetración de Internet y sus tecnologías en nuestros procesos, los cambios se sucedieron cada vez más rápidamente. A los registros bibliográficos legibles por ordenador siguieron lo que por entonces se denominó genéricamente como “recurso electrónico⁵”. Los catálogos comenzaron a incluir hiperenlaces a los textos descritos por esos mismos metadatos. Y esto supone otro cambio clave: entramos en un mundo en el que tenemos que gestionar no sólo información referencial sino también información de contenido. Se empiezan a enriquecer los catálogos con nuevas prestaciones y funcionalidades, fundamentalmente debido al potencial de la Web. El último hito de esta segunda fase ha sido la creación de los repositorios institucionales⁶ para asumir la responsabilidad de la gestión de esa información digital que se genera en una institución u organización, y que no tiene cabida ya en un sistema de gestión tradicional de información como es el catálogo.

Sin embargo este proceso de digitalización de los sistemas documentales sigue muy unido al soporte impreso. Se trata de una fase de emulación en el que la información y su tratamiento conservan las características de procesamiento de este tipo de soporte. Los objetos digitales reproducen el objeto impreso, así como las herramientas para su gestión. Los sistemas de información digital abarcan no sólo una tipología documental mucho más variada que en el mundo impreso, sino que incrementan la capacidad de acceso y visualización de la información por parte del usuario.

Uno de los elementos característicos del sistema de información digital que lo liga aún hoy en día al sistema impreso es el relativo al concepto de “incunable digital”⁷. Se trata de objetos digitales cuya forma permanece enraizada en la tradición impresa, con los formatos

⁴ Reconceptualización originada por el hecho de introducir un nuevo usuario en el sistema que es el ordenador. Se trata de un nuevo usuario con capacidad para generar una gran demanda de información, pero sin capacidad para evaluar la pertinencia y calidad de la misma.

⁵ Primero llegaron los libros y revistas electrónicas y luego las colecciones electrónicas que se iban incorporando como resultado del proceso de digitalización. RODRÍGUEZ BRAVO, B. *El documento: entre la tradición y la renovación*. Gijón : Trea, 2002.

⁶ ROVIRA FONTANALS, C. ; MARCOS, M.C. ; CODINA, L.I. Repositorios de publicaciones digitales de libre acceso en Europa: análisis y valoración de la accesibilidad, posicionamiento web y calidad del código. *El profesional de la información*, vol. 16, nº 1, 2007, pp. 24-25.

⁷ CRANE, G. What do you do with a million books? *D-Lib magazine*, vol.12, nº3. En: www.dlib.org/dlib/march06/crane/03crane.html. Consultado el 15/02/2010.

html y pdf mimetizando las limitaciones de sus predecesores impresos. Esto implica que las actuales herramientas de gestión de la información continúan apegadas al uso que se hacía de la información en la era impresa, construyendo silos de información estática, y cuyo principal valor añadido, respecto al mundo impreso, es que ahora es accesible remotamente desde un ordenador.

No obstante, el mundo digital se impone y el alcance de las transformaciones que la tecnología de la información en red está provocando en los procesos científicos y de asimilación y transmisión de contenidos es de tal magnitud que está cuestionando los conceptos y premisas hasta ahora imperantes en todas las ramas del conocimiento relacionadas con la creación, transmisión y custodia de contenidos⁸.

En el mundo de la información en red, no sólo cambian los escenarios, las actitudes y el uso de esa información, sino el propio concepto de lo que es un documento y por tanto la forma en que debemos gestionarlo para poder dar soporte a aquellos procesos en este contexto exclusivamente digital⁹. Por ello, el campo de la Documentación en particular, y de las Ciencias de la Información en general se encuentra inmerso en un momento de cuestionamiento de su identidad y de su papel en la sociedad.

En nuestra tesis planteamos que el documentalista debe dejar de cuestionarse cómo transformar los flujos de trabajo en su centro de documentación, para pasar a preguntarse cómo se están transformando los entornos de nuestros usuarios cuando trabajan en un entorno digital. Es en ese momento cuando el documentalista se plantea el modelo de gestión de la información como servicio de soporte y apoyo efectivo tanto a las demandas informativas de sus usuarios como a sus procesos de actividad en un entorno que unas veces es digital, otras multimedia y otras en papel.

El primer factor que está actuando como motor de esas transformaciones es la propia evolución de la web. Si aceptamos que la web es la plataforma universal para el uso y la comunicación de la información entonces debemos estar preparados para apoyar los procesos de información de nuestros usuarios sobre esa plataforma¹⁰.

⁸ CASTELLS, M. *La sociedad red*. Madrid : Alianza, 2005.

⁹ RODRÍGUEZ BRAVO, B. *El documento: entre la tradición y la renovación*. Gijón : Trea, 2002.

¹⁰ LAZAR, J. *Web usability. A user-centered design approach*. Boston : Pearson, 2005, p.6.

Una de las preguntas que nos hacemos los documentalistas sobre el sentido de nuestra profesión es acerca de la utilidad de nuestros servicios a la comunidad para la que trabajamos. En una época en la cual las tecnologías de la información tienen un papel protagonista en el desarrollo de la Humanidad, las ciencias de la Documentación y de la Información son complementarias y al mismo tiempo muy dependientes de otras ciencias más ligadas a la tecnología como la informática.

Creemos, como así esperamos demostrar en la presente tesis, que el profesional de la Documentación, concibe las nuevas tecnologías como herramientas de intermediación entre la organización del conocimiento humano y la realidad de los objetos físicos y abstractos, cuyo propósito es mejorar la eficiencia del hombre para generar nuevo conocimiento.

Veremos cómo el desarrollo futuro de la Documentación discurre ineludiblemente no sólo por el desarrollo de la Teoría de la Documentación y la Ciencia de la Información sino, sirviéndose del impulso que los avances de otras disciplinas le dispensan, construyendo estructuras de organización del conocimiento adecuadas a las propiedades intrínsecas de los documentos y de las nuevas exigencias informativas de sus usuarios.

Estos impulsos repercuten en el desarrollo de áreas de conocimiento específicas de la Documentación como la arquitectura de la información, la organización del conocimiento, el análisis de contenido, los lenguajes de indización y la interacción hombre-máquina o HCI. Esto produce desarrollos asimétricos que lastran el desarrollo global de las ciencias de la Documentación. Además supone una invitación a que otras ciencias, como las de la Computación, suplanten el acervo profesional de la Documentación. No obstante, las soluciones que a nivel teórico aportan esas otras disciplinas resultan muy deficientes en su propósito de proporcionar unos servicios eficientes de información y documentación al usuario. Una máquina o un programa difícilmente van a poder suplantar la labor de un profesional de la Documentación en el acercamiento del conocimiento al usuario, dada su labor intermediaria entre la sutileza de los contextos y los distintos perfiles de usuario.

En este sentido abogamos por un papel protagonista del documentalista en la selección de los sistemas de gestión documental que puedan complementar su labor intermediadora, pero en ningún caso podrá ser sustitutiva.

Es por esta razón que hemos recogido aquí una línea de investigación de mucha actualidad tanto en el campo de la Documentación como de la lingüística computacional como son los lenguajes de especialidad y concretamente el lenguaje jurídico, lo cual tiene ya más que ver con el campo del Derecho.

Respecto a la Informática hemos debido tratar con uno de sus campos de mayor desarrollo como es el diseño y aplicación de ontologías.

Sin embargo, ¿cuál es la idea que subyace a la indagación en todos estos campos?, ¿cuál ha sido el propósito de este trabajo y las metas que hemos perseguido alcanzar¹¹?

Toda investigación es conducida por una idea hipotética de la solución que pretendemos dar a una pregunta que nos inquieta. Nuestra hipótesis científica es una suposición, un enunciado teórico supuesto y no verificado, pero probable, referente a variables o a la relación entre las variables¹². Y, ¿cuál es la suposición de la que partimos? El hecho de que será posible el trabajo diario del documentalista en su biblioteca digital bajo la concepción de la “Web Semántica”. Una concepción completamente nueva, no sólo del documento web sino de toda la Documentación. Berners-Lee, al formular la Web Semántica la concibe como¹³:

La Web Semántica no es distinta de la Web tal y como la conocemos, sino que es una mera extensión, donde la información se proporciona con un significado bien definido y permitiendo un mejor trabajo cooperativo entre las máquinas y las personas. Los primeros pasos para tejer una Web Semántica a partir de la Web existente ya están en marcha. En el futuro, estos avances propiciarán nuevas funcionalidades para que las máquinas puedan llegar a procesar y “comprender” los datos que hoy día tan sólo pueden presentar y exponer.

Aún así no era consciente que su idea por prolongar la capacidad de la Web, para que no sólo los hombres pudieran comprender el significado de sus contenidos, sino que las

¹¹ Si bien la consecución de las mismas es algo que el tribunal de doctorado valorará en el correspondiente acto de lectura de tesis.

¹² SIERRA BRAVO, R. *Tesis Doctorales y trabajos de investigación científica*. Madrid : Thomson, 2005, p. 348.

¹³ BERNERS-LEE, T. ; HENDLER, J. ; LASSILA, O. The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American Magazine*, 2001. En: <http://www.sciam.com/article.cfm?id=the-semantic-web&print=true>. Consultado el 15/02/2010.

máquinas también sean capaces de hacerlo, estaba transformando la concepción del Documento, las formas de acceso a sus contenidos y por tanto los propios principios de la Documentación.

Por ello, nuestra hipótesis resulta tan estimulante como ambiciosa, pero no menos inquietante, especialmente en una profesión que se ha ido haciendo dependiente de la tecnología y de las disciplinas afines, a medida que el número de documentos digitales se multiplica de año en año.

Desde la máquina Memex de Vannevar Bush hasta el papel de la tecnología en los distintos modelos de comunicación, se ha contribuido a unificar la lectura y la escritura en un mismo soporte. El objeto de esta tesis ya fue en su día recogido por Bush cuando afirmaba que una de las características de su máquina Memex¹⁴ era no solamente su capacidad para recuperar la información y anotarla, sino también integrar un sistema de “índice por asociación”. Sistema que en el marco de la documentación hipertextual se denomina nexo¹⁵, “cuya idea básica es la capacidad de cualquier artículo para, a su vez, seleccionar, inmediata y automáticamente, otro artículo¹⁶”.

Pensamos que los topic maps proporcionan esta capacidad, a una supuesta máquina Memex, para que un documento sea capaz de seleccionar aquellos otros documentos semánticamente pertinentes a los contenidos expresados en el mismo. Es decir, los topic maps proveen, en dominios bien delimitados, al documento de hipertextualidad semántica. Esta se obtiene de un contexto surgido a partir de una red semántica de sintagmas con unidad significativa mínima¹⁷.

Partimos de la premisa básica que para la ciencia de la información, la búsqueda y localización de información contextualizada es su materia central. Otros problemas como la representación de los documentos o su descripción no son más que cuestiones derivadas de la primera. En buena parte, el análisis de la búsqueda de información se basa en la

¹⁴ BUSH, V. As we may think. *Atlantic Monthly*, 176, 1945. pp. 101-108. En: <http://www.theatlantic.com/doc/194507/bush> Consultado el 15/02/2010.

¹⁵ LANDOW, P. *Teoría del hipertexto*. Barcelona : Paidós, 1997.

¹⁶ Ibid. p.34.

¹⁷ MOREIRO GONZÁLEZ, J.A. *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural*. Gijón : Trea, 2004, p. 178.

psicología, lo cual es lógico si tenemos en cuenta que esta actividad se lleva a cabo a través de la navegación que los usuarios llevan a cabo entre los recursos informativos con independencia de su soporte.

Concretamente partiremos de la psicología cognitiva de Chomsky¹⁸, quien considera los estados conscientes del hombre estrechamente ligados a los constructos socioculturales en los que se encuentra. En este sentido, el diseño de los programas de recuperación de información no se lleva a cabo sin tener en cuenta herramientas de procesamiento del lenguaje natural, las técnicas de interacción hombre-máquina, o la semiótica. Así, en la recuperación semántica de información, ésta se efectúa por significados, por el sentido y la materia de los discursos. De manera especial cuando los documentos no son textuales como en imágenes fijas, imágenes en movimiento o documentos multimedia¹⁹.

La presente situación de sobreabundancia de información, producida en un intervalo de tiempo muy breve, ha hecho que las nuevas tecnologías de la información alcancen un papel muy destacable en la obtención de soluciones para la recuperación pertinente y precisa de la información y sobre todo para su organización con objeto de obtener nuevo conocimiento.

En una situación bastante análoga nos recuerda el papel de los primeros bibliotecarios como depositarios, custodios y organizadores del Saber Universal, los actuales bibliotecarios y documentalistas se ven obligados a transformar sus habilidades y experiencias de acuerdo con la transformación no sólo del concepto de biblioteca y de documento sino también de los hábitos de lectura y de las formas de adquisición del conocimiento.

Tratamos de explicar las características y posibilidades de una tecnología creada para la descripción de estructuras de conocimiento y su asociación con las fuentes de información. Esta tecnología es conocida como Topic maps. Los topicmaps son un metaíndice, cuya estructura, normalizada en ISO/IEC 13250:1999, permite la navegación a través de

¹⁸ CHOMSKY, N. *El lenguaje y el entendimiento*. Barcelona: Planeta-Agostini, 1980.

¹⁹ Es interesante en este sentido el trabajo del profesor Peter Enser, quien ha presentado los resultados de algunos proyectos de investigación recientemente acerca de las posibilidades de la recuperación de imagen basada en contenido, (CBIR) en relación con el papel que tendrá en el diseño de la Web Semántica. ENSER, P. ; SANDOM, C. ; HARE, J ; LEWIS, P. Facing the reality of semantic image retrieval. *Journal of Documentation*, vol. 63, nº4, 2007.

múltiples recursos de información de forma coherente. Los objetos de información asociados a un topic map pueden encontrarse en muy distintos formatos como HTML, PDF, formatos de imagen estática como JPG o TIFF y formatos de vídeo; pero el topic map, proporciona una configuración semántica que no es jerárquica²⁰ sino asociativa. Esto facilita la navegación entre los documentos con independencia de la forma, el formato y la ubicación de los mismos. Podríamos decir que esta tecnología está pensada para la localización de información ubicua porque partiendo de los sistemas de organización del conocimiento tradicionales como índices, tesauros, glosarios y redes semánticas pasa a estructuras organizativas más complejas, capaces de proporcionar un contexto a los documentos recuperados.

Por tanto, como veremos, la principal cualidad de esta tecnología es que permite trabajar a dos niveles:

- a) la colección de documentos,
- b) la organización del conocimiento, entendida como metaíndice en el cual sus términos apuntan a uno o más documentos de la colección.

Para una mejor comprensión sobre la aparición de esta tecnología se debe entender el proceso de crecimiento y desarrollo de los documentos electrónicos de modo similar a como se produjo el desarrollo de la documentación impresa a lo largo de la Historia de la Humanidad, desde las primeras tablillas en las que se hacían anotaciones contables de los recursos de un Estado, hasta la época Helenística en que se producen los primeros intentos de organización de los rollos de papiro y pergamino en la Biblioteca de Alejandría²¹.

Es con la difusión de la imprenta en los siglos XV, XVI y XVII cuando con la explosión y enorme difusión de los textos surge la necesidad de la recuperación de la información. Esta recuperación podía ser de dos tipos según se tratara bien de recuperar información dentro de una unidad coherente y continua de texto (en un documento como podía ser un libro) o

²⁰ Si bien es cierto que la manera de visualización podría ser así. En el capítulo correspondiente a la visualización de los topic maps veremos los tipos de visualización, y por tanto de navegación, para arquitecturas de información con topic maps. Más información acerca de los sistemas de visualización en topic maps lo tenemos en García Martul, D. ; Marzal, M.A. Visualización en Topic Maps: Tendencias y propuestas. En *Actas de las II Jornadas de Tratamiento y Recuperación de la Información*. Leganés, 2003.

²¹ ESCOLAR, H. Op.cit.

bien se tratara de recuperar la información en distintas unidades de texto (documentos). Es así como se crean los primeros instrumentos de recuperación, se crean los índices alfabéticos, de materias, de antropónimos, de topónimos; se crean los sumarios que recogen la división física del documento pero también su evolución explicativa. Surgen los primeros diccionarios que proporcionan no sólo una definición de los términos empleados en los documentos sino que facilita el acceso a documentos en otros idiomas y la recuperación de información con índices multilingües.

Por otra parte es en el siglo XVII cuando se desarrolla la corriente bibliófila, se hace patente la necesidad de recuperar todo lo publicado hasta el momento, se desea recoger en bibliografías exhaustivas todo lo publicado sobre una materia e incluso se quiere publicar una *Bibliographia Universalis* en la que se recoja todo lo publicado hasta el momento.

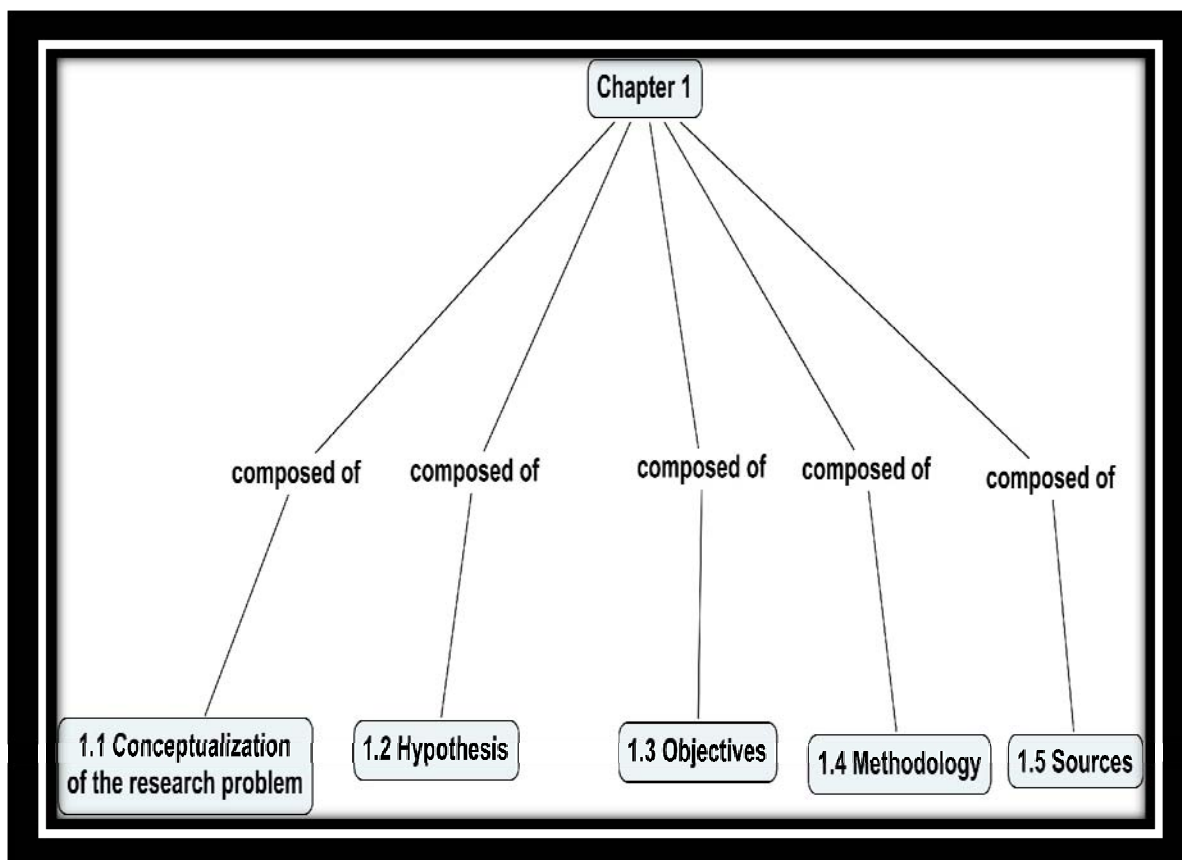
Es aquí donde debemos avanzar a nuestro tiempo presente, por cuanto nuestras primeras tablillas se han convertido en los formatos binarios de escritura más elementales, ya que la necesidad del hombre ya no es sólo recordar para sí mismo sino que debe enseñar a las máquinas a recordar. El formato de escritura para los ordenadores se va desarrollando a la vez que surgen distintos formatos y nuevas posibilidades de lectura-escritura como son la hipertextualidad. El surgimiento de la Web, el etiquetado (marcado) de documentos en un lenguaje común como es el HTML ha permitido la difusión global y universal de la información al igual que la difusión del griego en la cuenca del Mediterráneo permitió durante siglos la difusión del conocimiento en todo este entorno. Sin embargo será con la difusión del metalenguaje XML cuando se pueda integrar una gran variedad de formatos documentales ya que esta tecnología permite que sea el propio autor del documento quien cree sus propias etiquetas y por lo tanto hace que el ordenador lea el documento de la manera particular que el autor haya deseado.

En la actualidad estamos ante una nueva explosión de documentos, una revolución documental análoga a la acontecida con la aparición de la imprenta: el acceso a la información se facilita enormemente y es un acceso global porque cualquier lector puede acceder a cualquier documento desde cualquier punto del mundo sin necesidad de desplazamiento físico, no necesita tenerlo entre sus manos, pues se puede acceder a cualquier tipo de información desde cualquier parte del mundo en tiempo real. Los bibliófilos, los bibliotecarios de los siglos XVI y XVII pasan a ser los Bibliotecarios y

Documentalistas del mundo actual y comienzan a hacerse las mismas preguntas, ¿cómo se puede recuperar la información pertinente, la que interesa en cada momento sin perderse en una ya no inmensa biblioteca sino casi infinita? Surgen las mismas preguntas y como se expondrá en esta tesis se está respondiendo de la misma manera que en el pasado pero acordes con el hecho de que los documentos textuales han pasado a ser documentos digitales. Primera pregunta, ¿cómo se puede buscar una materia dentro de todos los infinitos “papiros” (páginas web) de nuestra biblioteca de Alejandría (Internet)? Pues en primer lugar se deberá contar con un bibliotecario que organice físicamente la biblioteca de forma que los papiros que hablan de una materia concreta estén situados físicamente en el mismo espacio. Sin embargo, hay un problema respecto a la organización de la información, porque los textos están muy divididos, demasiado segmentados y en muchas ocasiones repiten la información, ¿cómo se puede buscar de entre los miles de papiros que tratan de una materia, uno en particular y que proporcione la mejor de las respuestas posibles a nuestras preguntas sobre la materia? Es aquí cuando surgen los primeros libros, y junto con los libros están los índices que son a la vez el mapa del libro y el esqueleto que articula los distintos apartados. Análogamente hasta ahora en Internet se ha contado únicamente con hojas de libros mezcladas, pero el usuario necesita adquirir conocimientos sobre un tema en particular y para ello debe ser capaz de recuperar todas las hojas que comentan esa materia y luego organizarlas en un todo coherente; es por ello que surgen los topicmaps como un metaíndice que recupera los documentos pertinentes y los organiza y cohesiona de acuerdo con la inquietud cognoscitiva de cada usuario. Por lo tanto los topic maps serán el instrumento que cohesione los futuros libros digitales constituidos por páginas web y en el que el lector pasará a realizar una lectura más asociativa que jerárquica hasta el punto de que el lector de los documentos impresos pasará a ser un lector-editor de lo que lea.

CHAPTER 1.

Hypothesis, Objectives and Methodology



The disciplines constitute a control system in speech production, setting its limits through the action of an identity that adopts the shape of a permanent reactivation of rules. (Michael Foucault; 1970)

1.1 Conceptualization of the research problem.

When man travels through an unknown landscape, he tries to locate points of reference with which to orient himself in the space. We could describe this as an environmental appropriation task by means of spatial information control skills and techniques²². Analogously, information on the Web is a vast informative territory for which it becomes necessary to establish references with which we can orient ourselves, empowering us to assimilate information and thereby construct knowledge.

²² FOUCAULT, M. *La arqueología del saber*. Madrid: La Piqueta, 2002.

Maps are, therefore, an abstract representation of reality and can be understood as an abstraction that is executed from a particular point of view regarding a parcel of reality.

In accordance with this concept, the current search for information has a peremptory need of a normalized model. A model that guides the user through a network made up of information subjects in the satisfaction of its informative demands; as well as one that is capable of providing a framework with which to formulate different information recuperation and localization strategies.

This framework is made up of virtual information maps, adequate for keeping the vast informative territory available in digital format within the users reach, both on the Web and the invisible Internet; prepared not only to represent content but also a semantic context that guides the user.

We believe that topic maps constitute this virtual map that is representative not only of content but also the contexts in which it is inserted and that facilitate, both for people and machines, recuperation not only by information subject syntax, but also by its meaning.

Policy ISO/IEC 13250, which regulates the definition of topic maps, describes the elements that constitute this model and declares an exchange and interoperability format between topic maps.

A topic map, as we shall see, is basically a document with a particular syntax, derived from XML, named XTM, in which different types of elements are employed to represent topics, topic occurrences and the relationships or associations between topics.

A topic map expresses someone's opinion of what that person understands as topics and to which elements amongst a collection of informative resources each of them corresponds. Goldfarb was one of the first topic map theorists²³. He compared them with a GPS applied to the information space, where they are a basic framework for the representation and organization of knowledge structures.

²³ GOLDFARB, Ch. *XML handbook*. New Jersey : Prentice Hall, 2002.

The possibility of editing virtual information maps makes dealing with the original format of the information resources possible without having to carry out any type of conversion. This way a single resource can be used in many different ways thanks to different topic maps. By facilitating map transformation, the reuse of the information process is facilitated, as the topic map architecture is designed to fuse topic maps without requiring that the fused topic maps be modified.

Regulation ISO 13250, in its first edition defines a topic map by three characteristics:

A topic map is a collection of information sources used by an application as a collection of direct objects whose central document is a topic map document that adjusts to the SGML architecture defined by this international regulation.

Any topic map document that adjusts to the SGML architecture defined by this international regulation, or the document element (topic map) of said document.

The type of document element (topic map) of the topic map document architecture.

An information index that resides outside of the body of information can be created with topic maps. The topic map describes the information in the documents and the databases through its link with URIs²⁴.

The topic map takes the key concepts described in the database and documents and relates them to one another regardless of what the indexed documents affirm about them. The result is an information structure that breaks the traditional hierarchical model that up to now has been used for information extraction. In fact, a topic map contains many superimposed hierarchies that are enriched with associative semantic relationships. This makes the information much easier for the user to find, as it is he who constructs his own learning through the interrelationship of concepts, always keeping in mind that there are many avenues of consultation by which to reach the same response provided it is based on an appropriate concept for the completion of navigation.

²⁴ PEPPER, S. The TAO of Topic Maps. Finding the way in the age of infoglut. En: <http://www.ontopia.net/topicmaps/materials/tao.html>. Consulted on 12/04/2009.

The most common use of topic maps nowadays is for the construction of Web sites and their subsequent organization for navigation, locating the most pertinent information. Topic maps provide a structure, and the content of the page is taken partially from its topic map and partially from occurrences. This is an ideal solution for Web portals, catalogs and indexes. Additionally, while it can be said that a topic map represents knowledge regarding the things it describes, topic maps will be the ideal knowledge management tools.

1.2 Hypothesis.

The rapid development of information and communication technologies constantly modifies document properties, as well as their possibilities of interaction with users. Conventional knowledge organization systems have proven themselves incapable of taking advantage of all digital documentation properties. Specifically in the digital libraries of domains containing specific knowledge, such as legal knowledge, characterized by a specialized vocabulary and characteristic documentary patterns, there is an inappropriate use of thesauruses and taxonomies due to its inability to assist the user in the recuperation of legislation compared not only by states but also by different administrations. Additionally, if a language is used both pre-coordinately and post-coordinately, it does not facilitate the user's ability to use alternative search methods because of the limitations of the relationships established for documentary languages²⁵.

We believe that the appropriate organization of knowledge should be carried out by technologies that take into account the intrinsic characteristics of digital documents and the architecture of information specific to each digital library and its knowledge domain.

We believe in the technological-documentary framework of topic maps as they rely on a more flexible documentary metalanguage that is adaptable to the associative and hypertextual organization of knowledge, in which the user is an active protagonist in the recuperation of information, the organization of knowledge and the construction of his own understanding. Until now the traditional documentary languages, better adapted to the treatment of information on paper, with a predominantly hierarchical organization of its

²⁵ In this regard, the project of Professors Moreiro and Llorens, in the context of other international projects, has found a documentary language flexible enough to recuperate multimedia documents, for which they proposed a verb thesaurus. Moreiro González, J.A. El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural. Gijón : Trea, 2004, pp. 204 ss.

materials, have relegated the user to a passive role in the resolution of his informative demands.

We contend that the design of an information architecture, for the Center of Political and Constitutional Studies' digital library, based on a topic map with a specialized legal vocabulary, noticeably increases the user's experience in the navigation and localization of documents, regardless of their material, form, format and physical space, without losing the semantic context of the collection to which they belong. This is possible because the topic map model is based on a knowledge organization model created in the line of Documentation Sciences, but that was formalized by the Computation Sciences in regulation ISO 13250:1999 and its unique syntax, XTM. From this sprung a series of tools for the editing of topic maps applied to very different environments, from companies to public administration to semantic digital libraries²⁶.

We have centered the application of topic maps on the digital library portal of the Center for Political and Constitutional Studies, as it deals with a knowledge domain and a well-delimited vocabulary: the legal area and language. Additionally, the center has a recently created portal for its digital library, and it has automated its collections with Absys, which is poorly adapted to the consultation of digital collections.

1.3 Objectives.

The hypothesis proposed allows us to set a goal that we wish to attain with this thesis. The objectives specified represent the four cardinal points on which the thesis map is drawn.

We will define four general or cardinal objectives that will outline the general knowledge framework with which we will formulate our hypothesis regarding the pertinence of a general methodology for the design of topic maps; a methodology that will be tested in the creation of a topic map for the organization of informational resources in the Center for Political and Constitutional Studies' digital legal library.

²⁶ In the corresponding applications chapter, we will see some examples, such as the New Zealand digital library, a Call Center for the United States IRS and the City of Bergen's (Norway) Web site.

1.3.1 Cardinal Objectives.

- 1) **Determine** the characteristics of both the collection and the specialized language used in the documents that make up the document repository.
- 2) **Establish** value added services.
- 3) **Define** the role of information technology in the factual map.
- 4) **Design** an organization outline for digital legal documentation.

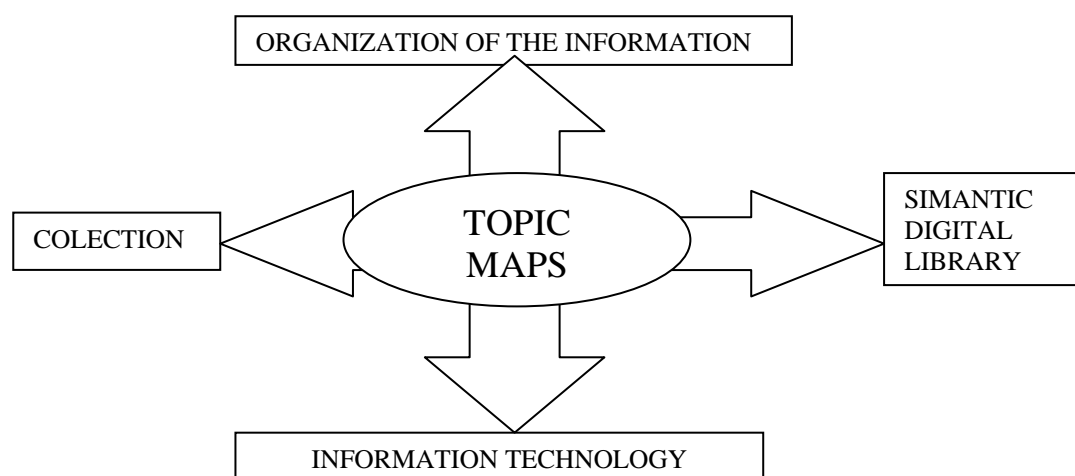


Figure 1. Chart of the interrelationship of topic maps in the thesis.

The following objectives are an intersection of three scientific frameworks as shown in the following graphic.

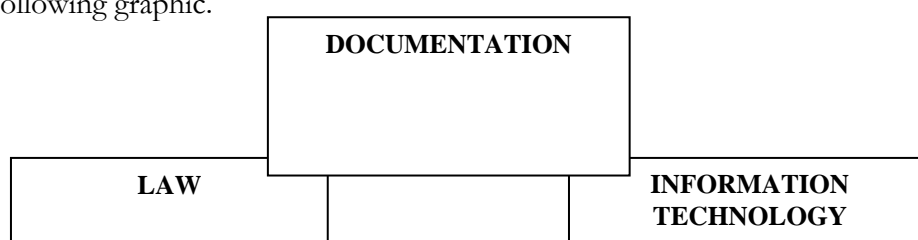


Figure 2. Chart of the interrelationship between the scientific fields involved in the development of the thesis.

The reason it has become necessary to deal simultaneously with these three frameworks is because of the intrinsic and peculiar characteristics of the center in which the proposed

methodology is going to be applied. Obviously, the ramifications of each one makes a detailed study and analysis of the three impossible.

Aspects of Law of such interest as specialized legal language, the role played by citizen Internet access to legal studies that the executive power mandates, and other issues more directly related with the processes of production and management of legal documents, are treated here in order to obtain the utmost scientific precision. Notwithstanding, some aspects must be treated superficially so as not to lose sight of the objective of this work: the formulation of a methodology for the application of topic maps to a legal digital library.

Similarly, the ramifications of documentation are, as we know, numerous. The methods and techniques employed by this science to analyze and describe documents, both externally and internally, are very numerous and complex, as reflected by the effect of the constant change in international regulations on the formal description of documents²⁷. In the specific case of analyzing document content, we expect a changeable reality with regard to the controlled languages.

The third disciplinary frame that supports this work is that which we refer to as Information Technology, equally a multitude of very interesting and important derivations for the treatment of legal information can be established. Subjects such as the relational or Chen database design models, information recuperation theories or ontology design should also be treated collaterally; concentrating on those IT derived technologies that are directly applicable to the area of this scientific work.

Attempting to join these three fields in a single scientific work means an inevitable loss of depth in many of the subsections considered interesting within each of them, but this facilitates the procurement of a highly enriching general idea of the model here proposed for the application of the nine proposed documentary languages applied to the analysis of the content in the digital environment.

For the exact delimitation of the partial objectives pursued in this doctoral thesis, it is necessary to clarify the importance that the practical component has had from the very

²⁷ New formats exist for the formal description of documents, such as FRBR or the new MARC formats, but there are also new languages for the description of content, such as SKOS.

beginning. The intention is to propose an investigative work that is not limited solely to portraying the updated succession of ideas, activities and achievements of other researchers, but rather to include our own experimental contributions. This means that from the first moment since the conception of this thesis, it has been perfectly clear that a practical development had to be proposed as an application of all of the ideas conceptually defined in this work. This fact has led to a double effort, as in addition to the compilation and consultation of sources and the study and synthesis of ideas explained by other researchers on the subject, there is the acquisition of abilities needed for the use of topic map technology techniques and tools.

1.3.2 Specific Objectives.

Propose a methodology for the application of a semantic knowledge recuperation technology and the organization of knowledge for a legal studies digital library, the Center for Political and Constitutional Studies (CEPC), entitled: Topic map.

Show a global vision of the regular function of the Center for Political and Constitutional Studies' library, with a focus on the users' use of the Web site and digital library, using a quantitative surveying methodology.

Prove, by means of a qualitative focus group methodology, the difficulties experienced by those in charge of the digital library and users when sharing materials and locating information, taking into account that the documentation being dealt with is found within a collection with a very specific domain of knowledge, with a specialized language and very specific documentary methods.

Place at the disposal of librarians and documentarians from legal training and research centers an innovative formula for how to work with documentation, based on the idea that it no longer makes sense to work physically with the documents, but rather that it is indispensable to work with content and its own specialized language.

1.4 Methodology.

1.4.1 Contextual Methodological Framework.

Computer science researchers have maintained two basic attitudes when approaching research²⁸: a) positivist, and b) phenomenological.

Positivism attempts to understand the causes of social phenomena, independent of the subjectivity of individuals²⁹. The phenomenological attitude attempts to understand those same phenomena from the point of view of the person who experiences them. Given that positivists and phenomenologists deal with different types of problems and search for a different class of response, they use different methodologies. While positivists adopt the natural sciences research model, using techniques, such as questionnaires, that generate quantitative data capable of being analyzed by statistical methods. Phenomenologists employ qualitative methods, such as participant observation and in depth interviews that obtain descriptive data³⁰.

Qualitative research³¹ is a research process that extracts data from the context in which phenomena take place, in an attempt to describe them, as a way to explain the process in which they are inserted and the different focus perspectives; employing induction in order to obtain all of the possible explanations for the foundation of the observed phenomena.

The paradigm of qualitative research versus quantitative research is that the significance of the phenomena and their interactions might only be understood by the subjects participating in a specific situation. The ultimate objective is to understand the facts analyzed in context. And although the researcher participates in the reality studied, the distortion that this can produce is less than with other research methods³².

Qualitative research is one of the principal research methods employed in the social sciences, along with the quantitative method, which is more focused on the quantitative assessment of the phenomenon's behavior but closer to the positivist paradigm of the

²⁸ PICKARD, A. J. *Research Methods in Information*. London: Facet, 2007, p.6

²⁹ Id.

³⁰ NEUMAN, D. High school student's use of databases: results of a national Delphi study. *Journal of the American Society for Information Science*, vol.46, n°4, pp. 284-298.

³¹ GORMAN, G.E. *Qualitative Research for the information professional*. London: Facet publishing, 2005.

³² Ibid., p.26

purist sciences with the collection and analysis of data, susceptible to reflecting observable incidences and measurable occurrences.

Qualitative investigation resides in the interpretivist paradigm, which is more centered on complex social constructs that because of their continuous evolution are more difficult to interpret in numeric terms.

Taylor and Bogdan list ten qualitative research characteristics³³:

Qualitative research is inductive. The concepts are developed based on data and not by collecting data through the evaluation of models or hypotheses. This way we can apply a more flexible research design to the development of a thesis.

The researcher sees the analysis scenario from a holistic perspective. The people, the scenarios and the groups are not reduced to variables, but rather considered something global. In our thesis regarding the semantic localization of an informative request resource, the users are studied under the context of their sociocultural environment and the difficulties that they encounter when resolving a problem.

Researchers are sensitive to the effects that they themselves generate for the user subjects of the study, and if they cannot eliminate their effect on the people they study, they try to reduce them. In the case of the participant observation that we have employed, we have tried not to ignore the phenomenon analyzed.

Researchers try to understand people in their basic reference environment. For the phenomenological perspective and for qualitative investigation, it is essential to capture the environment as the users experience it. Qualitative researchers identify themselves with the people they study in order to understand their perception of the context.

Qualitative researchers abate their own perspectives and predispositions. They watch the events as though they were happening for the first time, nothing is inferred.

³³ TAYLOR, S. J. ; BOGDAN, R. *Introducción a los métodos cualitativos de investigación*. Barcelona: Paidós, 1994, pp.23-24.

All of the perspectives that our users can contribute are valuable. We are not looking for the truth, but rather a detailed understanding of the network of perspectives that make up the perceptive context of a community of users.

Qualitative researchers highlight the internal validity of their research in relation to the trustworthiness and reproducibility of quantitative research.

Qualitative methods allow us to remain close to the empirical world. They are destined to ensure a close fit between the data and what the users really say and do. Observing users in their daily interaction with the library and the systems at their disposal, such as OPAC, databases, electronic magazines, the library Web site; a qualitative researcher obtains knowledge directly from the user, unfiltered by concepts, operational definitions and classification scales.

In qualitative methodology all scenarios and people are worthy of research. No aspect of the user's behavior is too frivolous or trivial to not be studied. All of the scenarios and users are similar yet unique. They are similar in the sense that in any scenario or amongst any group of people, general types of behavior can be found. They are unique because in each scenario or by means of each informant some aspect of the interaction with the library can be best studied.

Qualitative methodology has not been refined or homogenized as much as other research approaches. Qualitative research is flexible with regard to the way in which studies are conducted; it encourages the creation of unique methods, following orientation directives but not rules.

All of these qualitative research characteristics are more understandable if we compare them with the properties of quantitative research. They are compared in the following table created by Orozco Gómez, who explains the differences between the qualitative and quantitative perspectives in communication and computer sciences³⁴.

³⁴ OROZCO GÓMEZ, G. *La investigación en comunicación desde la perspectiva cualitativa*. La Plata: Facultad de Periodismo y Comunicación social, Universidad Nacional de la Plata; Guadalajara: Instituto Mexicano para el desarrollo comunitario, 1997.

Qualitative Perspective	Quantitative Perspective
Interpret	Objectivity
The distinctive	The average
Subjects	Events
Involvement	Neutrality
Premise	Hypothesis
Categories	Variables
Describe	Measure
Associate	Separate
Processes and results	Results
Methodological Creativity	Precise Techniques
Substantive Rationality	Instrumental Rationality
Study of microprocesses	Study of macroprocesses

Informed Theory

Statistics

Table 1. Differences between qualitative and quantitative methodologies.

The reasons why we have opted fundamentally for the qualitative methodology rather than the quantitative are³⁵:

We propose the need for a multiple conception of reality capable of condensing the factor complexity that intervenes in the semantic localization of information recourses. This requires the contemplation of phenomena in its execution context.

The main objective of this thesis is aimed at understanding the facts, which involves understanding of the possible latent and manifest relationships that underlie the satisfaction process of the users' informative requests during their library interaction.

The researcher and the research subjects are realities that are not separate but interrelated. This involves a series of relationships between the different study subjects that make up a map of mutual influence that must be attended to.

This allows us to develop a series of ideographic knowledge on the different individual cases studied, with the goal of isolating the generalizable characteristics that can be specific to a certain situation.

When the causes of an objective reality and its effects are unclear, qualitative research allows us to single out each of them.

We do not have to work with discrete values to infer a regular model of information recuperation among our users; it interests us to interpret the user behavior when encountering a new environment so as to improve the user's adaptation.

³⁵ GUBAN, E; LINCOLN, Y. Competing paradigms in qualitative research. En Denzin, N; Lincoln, Y. (eds.). *Handbook of qualitative research*. Thousand Oaks: Sage, 1994. pp. 105-117.

1.4.2 Qualitative methodology: characteristics and reasons for its use.

The choice of a methodology when conducting research involves a specific orientation in the development of the work, given that it indicates the important lines to follow as the study advances and the subsequent phases, whose correct culmination, give validity to the conclusions obtained. The variety in the methodology gamma is essential for the researcher, who can choose from amongst a wide selection of tools in order to pick the one that best adapts to the problems that he faces. In the library science field, however, there has been a traditional preference for a single paradigm, the positivist and the quantitative, to the detriment of another great reality analysis model, the phenomenological or qualitative³⁶ which we have favored in this thesis.

It is difficult to define “qualitative research” given the lack of its coherent use and the diversity of methodological focuses that claim to have a qualitative connection. It has gotten to the point where the validity of the label “qualitative research” is put into question. The progressive interest in qualitative methods in the social sciences has arrived somewhat late to the library sciences³⁷. Fidel observes the fact that research focuses more and more on the user as one of the reasons that has led researchers to notice qualitative methods. Other reasons behind this interest are, on the one hand, the failure of quantitative methods with regard to what was expected of them; for example, in information recuperation, a reductionist approach is taken by not taking into account individual situations and contexts that are essential to the recuperation and the subsequent organization of results. And on the other hand, the growing interest in qualitative methods in the social sciences³⁸.

This interest in phenomenological research is evident in publications of classic information management studies,³⁹ diverse manuals⁴⁰ on the application of qualitative methodologies

³⁶ SUTTON, B. The rationale for qualitative research: a review of principles and theoretical foundations. *Library Quarterly*, vol.63, n°4, 1993, pp. 411-430.

³⁷ BOGDAN, R. Foreword. En MELLON, C.A. *Naturalistic inquiry for Library science: methods and applications for research, evaluation, and teaching*. Nueva York: Greenwood Press, 1990.

³⁸ FIDEL, R. Op. cit.

³⁹ GLAZIER, J.D. *Qualitative research in information management*. Englewood, Colorado : Libraries Unlimited , 1992.

⁴⁰ GORMAN, G. E. *Qualitative research for the information professional : a practical handbook*. London : Library Association Publishing , 2005.

within library science, and different monographs in magazines such as *Library Quarterly*⁴¹, *Education for Information*⁴² and *Library Trends*⁴³ that today are classic references.

1.4.3 Qualitative methodology work in our thesis.

One of the characteristics of the qualitative methodology is its inductivism. In this thesis we do not attempt to collect information that allows us to verify or reject previously established hypotheses, but rather it is through the data collected that we try to reach an understanding of the phenomena studied in relation to the context in which it develops. This means gathering all of the variables, the context, the description, the process comprehension and the presentation of the different perspectives involved so that they can be analyzed jointly in a coherent and significant manner in a “less to more” approach once the data has been collected⁴⁴. We collect the evidence, and we use it to develop an explanation of the facts in order to establish a theory based on the phenomena observed. This is called the grounded theory, given that it is founded on the data found at the most basic level or is built up from zero, so that we begin with an inference and can interlace data to construct information.

The selection of participants was intentional. The subject sampling that participated in the study is very small in comparison with those of quantitative studies. The fact that the analysis is exhaustive compels us to strictly limit the number of participants, with the goal of ensuring the study's feasibility. However, in order to avoid excessive reductionism in the conclusions, we try to include as wide a variety of subjects as possible, always within quantitative limitations regarding the number of participants. In this thesis, we have conducted interviews with graduate students and library professionals. To some extent, working with informants of disparate characteristics, a difference which we hope to see reflected in the diversity of behavior, highlights that the research is not starting from zero, as the analysis process of the data does not respond to a strictly inductive model, but rather the preconceived ideas and expectations prompted by logic play an important role.

⁴¹ BRADLEY, J ; SUTTON, B. Qualitative research. [*Special Issue*]. *Library Quarterly*, vol.63, n°4,1993.

⁴² ZEITLYN, D. ; BEX, J. Libraries without number. [*Special issue*]. *Education for Information*, vol.15, n°4,1997.

⁴³ McCOMBS, G.M. ; MAYLONE, T.M. Qualitative research [*Special Issue*]. *Library Trends*, vol.46, n°4, 1998.

⁴⁴ GORMAN, G.E. Op. cit.

This resorting to the confrontation of information obtained from participants of disparate characteristics allows us to maximize the variety of data collected within a relatively limited field due to the quantitative limitation of the number of subjects studied. It is what on occasion is referred to as “theoretical or intentional sampling” as opposed to the “probability sampling” of quantitative studies⁴⁵.

On the other hand, the use of a research design with a qualitative methodology has provided us with flexibility when developing this thesis. The collection of data is intimately linked to its analysis, which requires being open to whatever may arise during the development of the study; and to be open to whatever may happen, the design must be flexible so that it can progress with the development of the thesis. More than analyzing the data once collected, it is the analysis of the different possibilities that the topic maps offer for the management of the semantic digital library that has guided the collection.

It is our interest in studying specific situations of qualitative research, as is the case with the CEPC digital library, that has incited us to its use, but this does not rule out the possibility of inductively establishing generalizations or developing abstractions. By contrast, although quantitative techniques are very useful to substantiating clearly articulated theories, the methodological restrictions imposed by quantification make the use of these methods for the generation of a new theory difficult. The methods most open to unexpected relationships are generally more productive when suggesting new theory, although so much information can be accumulated that analysis becomes one of the principal problems⁴⁶.

We have carried out the analysis of the data based on its categorization, in that we have studied the information collected and tried to establish nexus and links amongst the user contributions; data grouped by similitude in order to elaborate abstractions.

The use of the qualitative methodology requires a combination of several sources and data collection techniques with the goal of obtaining an in-depth understanding of the phenomenon studied, ensuring the credibility of the data collected, and therefore the validity of the conclusions. We have analyzed behavior in the use of thesauruses and topic

⁴⁵ PICKARD, A. J. *Research methods in information*. London : Facet Publishing, 2007, p. 239 ss.

⁴⁶ *Ibid.*, p.18.

maps amongst different users, not to establish comparisons amongst them⁴⁷ but to reduce the risk of omission of any type of behavior that varies depending on the user profile. This same way, the combination of various information collection techniques: open interviews and questionnaires have allowed us to contrast user behavior with controlled languages and topic maps.

The results of applying a qualitative methodology follow a narrative structure. From a formal point of view, qualitative research, like quantitative research, reflect results but present themselves as representations of the researcher's observations, especially when trying to maintain the qualitative character of the data. In these cases, there are narrative descriptions of the observations that we have experienced at the CEPC. This formal presentation sometimes makes it difficult to show the relationship between the collected data and the abstractions elaborated based on this data. To resolve this problem, we use techniques such as employing typical examples⁴⁸, which allows us to insert in the narration quoted examples from the information provided by the subjects and that help us illustrate our postulations.

Qualitative methodology is especially useful in the study of human conduct. It is the methodology best adapted to the exploration of user behavior when dealing with relatively unknown resources and tools⁴⁹ such as thesauruses and above all topic maps. For this reason, the data sources are always people, whether directly through interviews and observation, or indirectly examining quotes. The fact that qualitative methodology has been incorporated in the library sciences from disciplines such as sociology shows that there is a series of methods especially suitable for the study of people's conduct, whether its head to head with an automated bibliographic information recuperation system such as OPAC, when searching for information by navigating digital texts or dealing with a librarian. In general, we should say that the user's conduct is complex and is influenced by many factors that are both intrinsic to the user's personality and created by the environment, which makes it difficult to outline a group of variables that explain their conduct and whose manipulation always generates the same results. And it is not solely about the quantity and complexity of the variables involved, but rather these variables are not always the same in

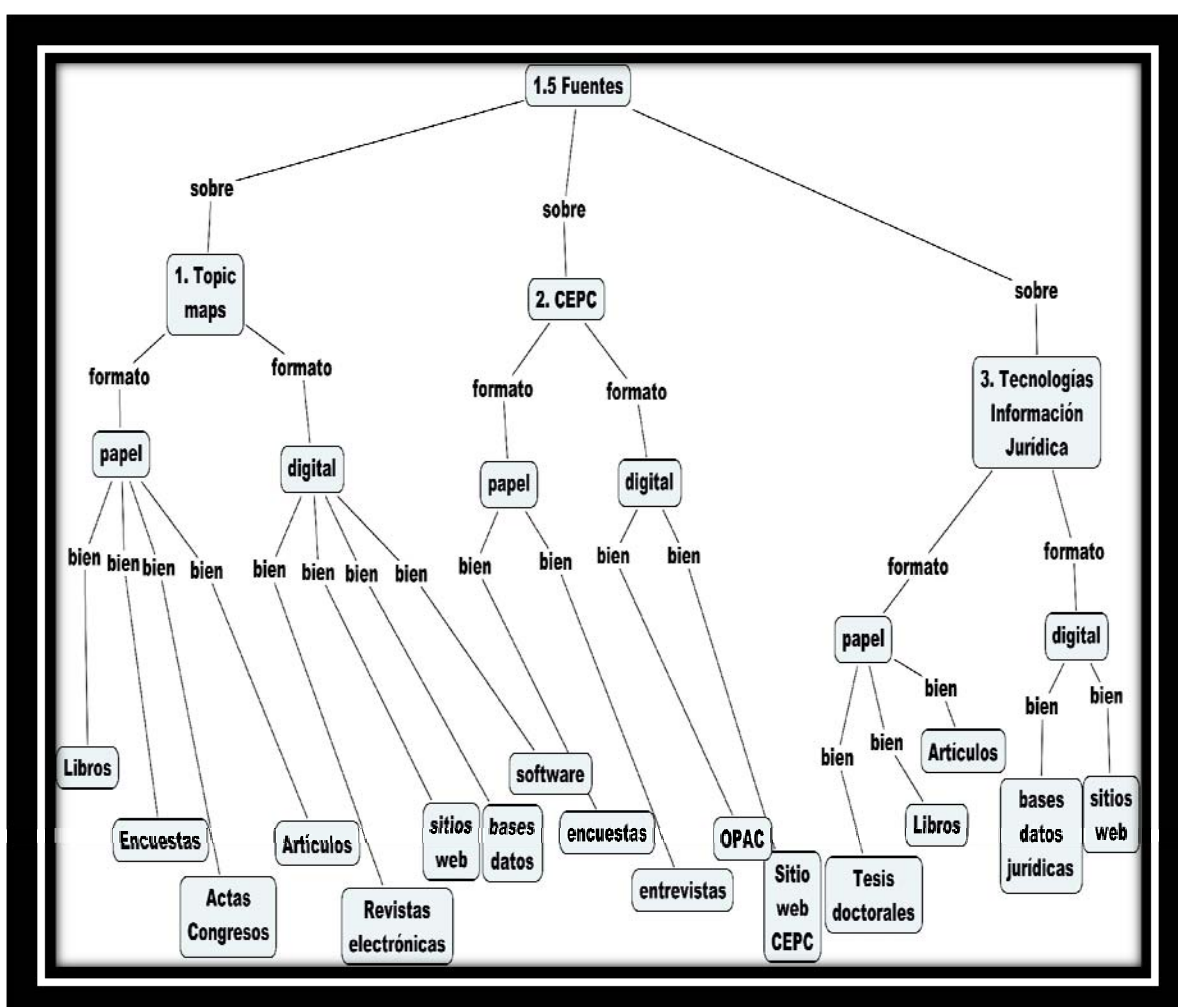
⁴⁷ Which would be impossible given the limited size of our sample.

⁴⁸ BRADLEY, J. Methodological issues and practices in qualitative research. *Library Quarterly*, vol.63, nº4, 1993, pp.431-449.

⁴⁹ FIDEL, R. Op. cit.

all of the individuals. This diversity of factors means that the best approach is not the analysis of a sample with the intention of extrapolating the results to the group because we would either leave out factors or the conclusions would be excessively vague. It is more efficient to focus on the exhaustive analysis of a reduced group of subjects, even knowing that the results cannot be generalized outside of the group of individuals studied, in order to reach an in-depth understanding of the reasons that the subjects are compelled to act as they do.

1.5 Fuentes.



Las fuentes que se han empleado para el desarrollo de la tesis son muy diversas y complejas a causa del planteamiento de una investigación teórica con propósitos prácticos.

Como podemos observar en el mapa conceptual siguiente, el ámbito de la investigación aglutina tres campos estrechamente interrelacionados:

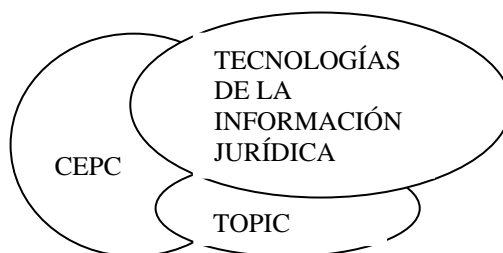


Figura 3. Interrelación de campos informativos empleados.

En el marco del Centro de Estudios Políticos y Constitucionales (CEPC) hemos tratado de aplicar una base conceptual sobre el significado de las tecnologías de la información jurídica y cuál es el papel que en las mismas tienen los topic maps.

Por ello, para hablar de las fuentes empleadas debemos hacerlo desglosando las fuentes de cada uno de estos campos, dado el carácter diverso de las mismas.

Por un lado las fuentes empleadas para la investigación acerca de los topic maps. No existe una extensa bibliografía en castellano acerca de este nuevo campo. Efectuando una búsqueda en el motor de búsqueda Google, si efectuamos una búsqueda de referencias a “topic maps” en español y en España nos arroja un resultado de 428000 registros, pero muchos de ellos son meras referencias a obras publicadas en inglés acerca de los topic maps, e incluso documentos que nada tienen que ver con los topic maps. En el motor de búsqueda semántico Swoogle se encuentran únicamente 8 ontologías referidas a ello. Un recurso más adecuado sobre bibliografía en castellano, Dialnet, localiza 13 documentos, de los cuales sólo uno es anterior al 2005.

En la base de datos TESEO de tesis doctorales no se hace ninguna mención explícita a los topic maps en el título en una sola tesis. Únicamente tenemos constancia de una sola tesis en la que se haya hecho referencia a los topic maps⁵⁰, si bien de manera muy escueta y sencilla.

Las referencias bibliográficas en papel y por los cauces bibliográficos convencionales sobre la materia resultan en general escasas, especialmente si se trata de bibliografía en castellano.

⁵⁰ Tesis doctoral de SILIÓ, T. La isomorfía lingüística sobre la base de la iteración recursiva, la coherencia estructural y la autosemejanza del lenguaje. Director: Tomás Albadalejo Mayordomo. Universidad Autónoma de Madrid, 2004.

No existe ninguna monografía ni capítulo de libro en castellano, y únicamente existen unos pocos artículos en revistas españolas, en una revista brasileña, otra mexicana, una más argentina y otra colombiana. Todos ellos publicados en revistas de calidad diversa, y en donde se realizan explicaciones sencillas del significado de los topic maps pero sin aportar nada a su desarrollo ni se exponen experiencias de aplicación.

Respecto a la calidad de las revistas en las cuales se han publicado artículos en España, merecen destacarse cuatro, dos en la revista electrónica Hipertext y dos en la revista El profesional de la información, si bien son mucho más rigurosos y citados los dos artículos de la revista digital dirigida por Lluís Codina, Hipertext, ya que, los autores de los artículos exponen los resultados de sus investigaciones en el contexto de un grupo de investigación que a comienzos del año 2000 comenzó a investigar sobre el potencial de los topic maps en recuperación de información multimedia⁵¹. Por el contrario los artículos publicados en El Profesional de la Información son más bien una noticia acerca del desarrollo y características de los topic maps⁵² y la reseña de un libro. Si bien los autores de la reseña⁵³ han presentado recientemente una tesis⁵⁴ que recoge el papel de los topic maps en el campo de la Informática y la Documentación.

Entre los grupos de investigación dedicados en España a los topic maps merece destacarse el formado por los profesores Moreiro y Llorens, quienes en el marco de un proyecto de I+D, que tenía por objeto indagar en la posibilidad de incrementar y sistematizar el empleo de relaciones asociativas para la recuperación de información multimedia y la reutilización de software, evaluaron el potencial de una norma ISO surgida precisamente con una sintaxis propia de los documentos multimedia, HyTime, lo suficientemente flexible para

⁵¹ MOREIRO, J. A. ; SÁNCHEZ CUADRADO, S. ; MORATO LARA, J. Panorámica y tendencias en topic maps. *Hipertext.net*, nº1, Mayo 2003. En: <http://www.hipertext.net/web/pag229.htm> Consultado el 15/12/2009.

⁵² EITO BRUN, R. Topic Maps y la indización de recursos electrónicos en la web. *El profesional de la información*, vol.12, nº 2, 2003.

⁵³ TRAMULLAS, J. ; GARRIDO PICAZO, P. Xml Topic Maps. Creating and using Topic Maps for the web. *El profesional de la información*, vol.14, nº 2, 2005, pp.156-159.

⁵⁴ GARRIDO PICAZO, P. El procesamiento automático de Documentación textual con información histórica: una aplicación XTM y DITA. Director: Jesús Tramullas y Antonio Hernández Pérez. Universidad Carlos III de Madrid, Madrid, 2008.

poder emplear una estructura amplia de relaciones asociativas en la recuperación de cualquier tipo de objeto de información⁵⁵.

En el ámbito internacional debemos destacar en primer lugar el origen estadounidense de este tipo de estructura, lo que supone que una gran mayoría de la bibliografía se encuentre en lengua inglesa. En segundo lugar, se originó en el ámbito de las iniciativas de lenguajes SGML para la recuperación de guías de software, lo que justifica que el formato de esta documentación sea fundamentalmente digital y accesible por web. En tercer lugar, al tratarse de una norma en desarrollo, buena parte de su literatura se corresponde con comunicaciones en congresos del ámbito de lenguajes de marcado como XML o los congresos de Markup Languages que tienen lugar todos los años en Montreal. En cuarto lugar, al tratarse de una norma ISO, existe mucha documentación en foros de discusión creados ex profeso sobre la materia, y documentos resultado de la actividad normativa del subcomité 34 de la organización internacional de normalización, así como documentos de consorcios de empresas de informática como OASIS creados para impulsar los lenguajes de marcado. Finalmente, existe bastante documentación en las páginas web de empresas ligadas a la venta de servicios de organización del conocimiento y gestión de la información como INFOLOOM, MONDECA u ONTOPIA⁵⁶.

Respecto a la bibliografía académica internacional debemos decir que no es un tema que haya sido objeto de atención más que puntualmente por parte de determinados grupos de investigación de los que se ha hecho alguna tesis en el campo de la informática en la que se ha tratado el papel de los topic maps para la recuperación y reutilización de software. Una de ellas es una tesis doctoral presentada en el año 2000 en la Universidad de Minho (Portugal). No obstante uno de los primeros trabajos académicos acerca de los topic maps se presentó en el campo de la lingüística computacional en la Universidad de Göteborg en el año 2002 por Anna Nordborg. En ella se explica el marco teórico de los topic maps como evolución de los índices, para pasar a continuación a explicar algunos de los principales elementos de la normas ISO que regulan la edición de topic maps. Lo original de este trabajo es que se trata de una de las primeras propuestas de topic map para un

⁵⁵ Más adelante explicaremos el sentido del objeto de información en Documentación pero por lo pronto lo entendemos de acuerdo con la explicación de Brophy, P. *The library in the twenty-first century*, 2007, p.181: *Cualquier objeto, incluyendo libros y revistas físicas, documentos electrónicos y multimedia que puedan formar parte de la colección de una biblioteca.*

⁵⁶ Empresa noruega absorbida en 2007 por otra empresa noruega de mayor entidad en el mundo de la gestión de servicios de información como es Bouvet.

corpus documental, del dominio de Farmacia, y por otro lado se evalúa el potencial del topic map como clasificador automático sobre un corpus documental de un dominio de conocimiento definido.

Respecto a los manuales empleados se debe decir que únicamente se ha podido consultar dos, uno de ellos publicado del año 2002, es una obra colectiva en la que los principales autores de topic maps exponen las distintas facetas de éstos; el otro manual se publicó en Internet en el presente año 2003 y se trata de una obra genérica pero muy clara sobre el significado de los topic maps, como su propio título indica “Manual de Topic maps” es una obra muy adecuada para quien comienza a interesarse por este tipo de lenguaje documental.

En lo referente a publicaciones sobre normalización de topic maps, se ha consultado en la página web de la organización internacional para la normalización las normas ISO/IEC, concretamente la del subcomité 34 del grupo de trabajo 3 (W3) sobre lenguajes de marcado, relativas a topic maps tales como la 13250: 2000 y su segunda edición del año 2002. También se han consultado las normas ISO 19756 (Lenguaje plantilla para topic maps), y la ISO 18048 (Lenguaje de interrogación para topic maps). En la página web de topic maps.org se puede acceder al lenguaje de marcado creado específicamente para topic maps llamado XTM.

Respecto a páginas web relativas al trabajo con topic maps se han consultado las webs de empresas tales como Ontopia, Empolis, Techila, Mondeca o Coolheads; de las cuales se debe destacar las demos como Opera (Ontopia) o las actas del Congreso XML 2002 Baltimore (Empolis), así como las herramientas informáticas que ofertan para la construcción de topic maps como el Omnigator de Ontopia.

Asimismo se ha hecho uso de las Actas del Congreso de XML Europa 2003 que tuvo lugar en Londres, y al cual asistió el autor de la presente tesis, de donde se han extraído algunos ejemplos de aplicabilidad de los topic maps para la gestión de centros de documentación tales como su uso en la gestión de la información en un centro de consultas de la agencia tributaria estadounidense.

Finalmente, en cuanto a artículos, debe destacarse el trabajo ya clásico de Steve Pepper “The TAO of Topic maps” que puede consultarse bien en la página web de XML Europa

2000 o bien en la Web de la empresa Ontopia junto con otros muchos artículos sobre la materia. Se debe destacar los artículos muy didácticos que se nos presentan en la página web de EasyTopic maps.

1.5.1 Fuentes de información general sobre Topic maps.

A) Introducciones

El artículo de Steve Pepper “The TAO of Topic maps: finding the way in the age of infoglut” proporciona una introducción libre de un lenguaje técnico acerca del significado de los topic maps en relación con campos y elementos de la Documentación como glosarios, índices, indización, gestión del conocimiento, representación del conocimiento, navegación, recuperación, redes semánticas, tesauros. Se explican los componentes de los topic maps: T (*topics*), A (*associations*) y O (*occurrences*) con sus conceptos derivados; relacionándolos con aquellos conceptos propios del ámbito de la edición y gestión de información, a la vez que trata de indicar algunos de los usos que puedan tener en el futuro.

El trabajo de Holger Rath “Topic maps Handbook” en donde se muestra el papel que los topic maps juegan en el campo de la gestión del conocimiento, cómo sería el diseño de una aplicación topic map, y cómo éstos pueden ser generados. En general es una introducción detallada a esta tecnología con numerosos ejemplos.

El trabajo de Pepper: “Euler, Topic maps, and Revolution”, fue publicado en las actas del congreso internacional XML Europe de 1999. Proporciona una de las primeras introducciones a la primera norma ISO 13250, ilustrando las áreas potenciales de aplicación, especialmente en relación con la edición electrónica.

En el congreso de XML Europa del año 2000 Colin Baird presentó una comunicación bajo el título: “Topic map cartography: a discussion of Topic map authoring” en donde defiende el papel de los topic maps como catalizador de la organización en colecciones de objetos de información destinada a promover la navegación multinivel por las mismas.

Asimismo en este mismo congreso Steve Newcomb y Biezunski presentaron el trabajo “Topic maps go XML” en donde analizan las posibilidades de expresión de la norma ISO 13250 con sintaxis XML, la creciente utilidad de los Topic maps en contextos web y la

labor del consorcio Topic maps.Org por conseguir una especificación de los topic maps en XML, XTM, para aumentar la difusión de los topic maps en los mercados de la edición y publicación web.

Finalmente como introducción cabría destacar el trabajo de Biezunski: “The XTM Guide: A beginner’s guide to the XTM syntax”. El propósito de esta guía es proporcionar a los desarrolladores web unas pautas en XTM para el desarrollo de aplicaciones basadas en topic maps. Para ello proporciona una DTD con sintaxis XTM comentada.

Biezunski, uno de los iniciadores de los topic maps presentó uno de los primeros trabajos sobre el estado de la cuestión en el desarrollo de los topic maps. En una comunicación presentada en el congreso Extreme Markup, “Specializing Occurrences in Topic maps by Association Template Subclassing”⁵⁷, presenta algunos de los avances en la investigación y aplicación de los topic maps, tanto a nivel de normalización como a nivel de características de los elementos que constituyen un topic map, incidiendo por primera vez en el papel de las plantillas para el elemento association. Posteriormente Freese presentó en ese mismo congreso “So why aren't Topic maps ruling the world?”⁵⁸ Allí se exponen problemas y preguntas orientadas a suscitar debate acerca de las posibilidades de los topic maps para generar la web semántica frente a otras herramientas como los metadatos o RDF.

Newcomb, quien comenzó a trabajar con topic maps junto con Biezunski en su empresa INFOLOOM, presentó una metodología para la construcción de web semántica sobre la bases de los topic maps, “A semantic integration methodology”⁵⁹. Según él la clave del problema de la integración semántica está en averiguar cuando dos topics son acerca de la misma materia, problema de la correferencia. Es un problema habitual en la recuperación cuando se trata de seleccionar los documentos recuperados como consecuencia de una búsqueda en distintas fuentes. La dificultad de este problema reside en que algunos topics

⁵⁷ BIEZUNSKI, M. ; NEWCOMB, S. En:

<http://www.idealliance.org/papers/extreme/Proceedings/html/2001/Biezunski01/EML2001Biezunski01.html> . Consultado el 15/12/2009.

⁵⁸ FREESE, E. *So why aren't Topic maps ruling the world?* En:

<http://www.idealliance.org/papers/extreme/Proceedings/html/2002/Freese01/EML2002Freese01.html> . Consultado el 23/12/2009.

⁵⁹ NEWCOMB, S. *A semantic integration methodology*. En:

<http://www.idealliance.org/papers/extreme/Proceedings/html/2003/Newcomb01/EML2003Newcomb01.html> . Consultado el 23/12/2009.

definen las materias acerca de los que se habla y otros topics pueden tener una relación tangencial con una materia.

En este sentido Vatant planteó la posibilidad de resolver esta dificultad en “Tools for semantic interoperability: hsubjects”⁶⁰, indicando las posibilidades de la interoperabilidad entre los distintos tipos de materias.

En el área de la gestión del conocimiento merece la pena destacar el capítulo de libro de Sigel⁶¹ “Towards knowledge organization with Topic maps”, donde se plantea la cuestión de ¿cómo podrían los servicios basados en topic maps aliviar la presión sobre el problema de la gestión del conocimiento acerca de la reorganización, mejora e integración semántica de documentos sobre materias heterogéneas? Se muestra cómo reaparecen los clásicos desafíos de la organización del conocimiento y cómo pueden ayudar los topic maps a resolver algunas de las cuestiones planteadas en este campo.

Pepper en el capítulo del manual de XML “Topic maps at work”⁶² nos proporciona una introducción a la norma topic map e ilustra algunas de sus aplicaciones potenciales, particularmente en los campos de la edición de enciclopedias y documentación técnica, y explica los requisitos necesarios para toda herramienta de edición de topic maps. En un trabajo posterior publicaría uno de los artículos más conocidos acerca de la norma ISO 13250, “Navigating haystacks and discovering needles: Introducing the new topic map standard”. Proporciona una introducción a la norma topic map (ISO/IEC 13250), en particular al dominio de la edición de enciclopedias. En él se discute acerca de la relación entre topic maps y la recomendación RDF del consorcio W3C.

“Ferrets and Topic maps: Knowledge Engineering for an Analytical Engine” de Mason acerca del motor analítico “Ferret”, desarrollado originalmente por el departamento de energía de Estados Unidos para la búsqueda de datos clasificados y de asociaciones en documentos. Presenta los resultados de las búsquedas de acuerdo con unas reglas formales,

⁶⁰ VATANT, B. *The Wheel and the hub*. En: <http://www.mondeca.com/lab/bernard/hsubjects.pdf>. Consultado el 23/06/2008.

⁶¹ SIGEL, S. *Towards knowledge organization with Topic maps*. En Park, J. (ed.) *XML Topic maps: Creating and Using Topic maps for the Web*. Boston : Addison-Wesley, 2003. En: <http://www.gca.org/papers/xml europe2000/pdf/s22-02.pdf>. Consultado el 23/06/2009.

⁶² RATH, H. ; PEPPER, S. *Topic maps at work*. En GOLDFARB, C.F. ; PRESCOND, P. (eds.). *XML handbook*. Englewood Cliffs, NJ: Prentice Hall,1999.

para lo que necesita de la existencia de una base de información estructurada que represente no sólo hechos individuales sino de una serie de reglas. Los sistemas de organización del conocimiento están evolucionando hacia formas que mejoren su portabilidad. Los primeros desarrolladores entendieron que la base de conocimiento se puede construir en XML por una serie de árboles jerárquicos que representan taxonomías. Sobre ellas se aplica un topic map para expresar las relaciones entre las taxonomías. El topic map se construye a partir de las reglas de clasificación y luego se aplica a los niveles más específicos, el nivel de recursos de información. De esta forma toda la base de conocimiento pasa a estar organizada de acuerdo a un topic map.

El informe “K-Discovery: Identification of Distributed Knowledge Structures in a Process-oriented Groupware Environment”⁶³ de Smolnik y Nastansky trata de los resultados del proyecto K-Discovery para la aplicación de topic maps a entornos basados en groupware. Los escenarios de entornos basados en groupware muestran problemas de acceso a estructuras de conocimiento en general y de estructuras de conocimiento organizacional en particular. Por medio de la aplicación de topic maps a las bases de conocimiento organizacional basado en groupware es posible cerrar el hueco entre conocimiento e información.

Posteriormente desarrollaron el trabajo en una comunicación presentada en el 35th Hawaii International Conference on System Sciences de 2002 bajo el título: “K-Discovery: Using Topic maps to identify Distributed Knowledge Structures in Groupware based Organizational Memories”⁶⁴. En él defienden que muchas organizaciones tienen muy integrado los sistemas groupware en su infraestructura de tecnologías de la información. Las bases de datos compartidas de estos sistemas forman memorias organizativas que comprenden todo el conocimiento de una organización recogido durante todo el periodo de su existencia. Un problema clave es cómo localizar la información relevante en una colección informativa en constante crecimiento y que además se encuentra distribuida en muy distintas bases de datos y repositorios. Las funcionalidades básicas en sistemas

⁶³ SMOLNIK, S. ; NASTANSKY, L. K-Discovery: Identification of Distributed Knowledge Structures in a Process-oriented Groupware Environment. En *Knowledge Technologies Conference* 2001. En: <http://www2.gca.org/knowledgetechnologies/2001/proceedings/Smolnik%20Paper.pdf>. Consultado el 15/12/2009.

⁶⁴ SMOLNIK, S. ; NASTANSKY, L. K-Discovery: Using Topic maps to identify Distributed Knowledge Structures in Groupware based Organizational Memories. En *Proceedings of the 35th Hawaii International Conference on System Sciences*, 2002. En: <http://www2.computer.org/portal/web/csd/abs/proceedings/hicss/2002/1435/04/14350106babs.htm>. Consultado el 15/12/2009.

groupware no son suficientes para apoyar a los usuarios en la localización del conocimiento requerido. Los topic maps proporcionan una base conceptual para la estructuración semántica de una red de hipervínculos adecuada a la navegación por una colección distribuida y en constante expansión. El proyecto K-Discovery introduce un marco conceptual, una arquitectura para la creación de estructuras de conocimiento por medio de la edición de topic maps en repositorios informativos.

Otro trabajo interesante acerca del empleo de topic maps para la organización de entornos informativos es: “Managing complex environments with Topic maps”⁶⁵ de Vatant. Aquí se propone una metodología general para la construcción de topic maps y su gestión en entornos de información compleja que implican una gran diversidad de recursos, objetos, conceptos y usuarios. La metodología se centra en la distinción entre el nivel de objetos y recursos de información y el nivel conceptual. Se proponen métodos de estructuración, interrelación y gestión de los distintos niveles implicados en la gestión del conocimiento con topic maps.

Más concreto y desarrollado se encuentra este artículo de Vatant: “Building Knowledge Universe”⁶⁶. Se propone un marco metodológico para la definición de topics y associations. Se centra en la distinción que debe realizarse entre información objetiva, universalmente compartible de objetos y recursos y la representación del conocimiento ligado al contexto de conceptos abstractos.

En el Congreso Europeo de XML de 2001, Marc de Graauw presentó una comunicación que bajo el título: “Business Maps: Topic maps Go B2B”⁶⁷, defendía la necesidad de contar con herramientas capaces de mejorar las habilidades de los usuarios en tareas B2B. Para ello se necesitaría contar con estructuras de conocimiento portables, reutilizables y normalizadas. Es por ello que aboga por los topic maps como el sistema ideal para el diseño de “Business Maps”.

⁶⁵ VATANT, B. Managing complex environments with Topic maps. En *Knowledge Technologies Conference*, 2001. En: www2.gca.org/knowledgetechnologies/2001/proceedings/Vatant%20Paper.doc. Consultado el 15/12/2009.

⁶⁶ VATANT, B. *Building Knowledge Universe*. En: http://www.mondeca.com/index.php/en/r_d/publications. Consultado el 15/12/2009

⁶⁷ GRAAUW, Marc de. Business Maps: Topic maps go B2B. En *XML Europe 2001*. En: <http://www.xml.com/lpt/a/1013> . Consultado el 15/12/2009.

Posteriormente, Biezunski, en ese mismo congreso pero en 2002, desde su experiencia en gestión de información en el seno de su empresa INFOLOOM, exponía unos primeros resultados en la aplicación de los topic maps. Experiencia que recoge en su artículo: “The gap between structured and unstructures needs a bridge⁶⁸”. Título muy significativo del problema organizativo de la información digital, en donde declara que se necesita un nuevo puente que conecte la información estructurada con la desestructurada. La ineludible necesidad de localizar cualquier información es posible por medio de las materias de las que trata, sin tener que contar con la localización física o el formato de cada documento. Los topic maps reclaman un papel intermediador entre la información estructurada y desestructurada.

En este mismo congreso de XML 2002 Nishikawa, en su comunicación: “Organizing Information in a Corporate Intranet: A use case for published and internal-use subjects in topic maps⁶⁹”, defiende que el paradigma de los topic maps permite a los usuarios y organizaciones emplear una parte específica de la información direccionable, más conocida por “indicador de materia publicada”, como información vinculante acerca de una materia específica. Así, la información acerca de las distintas materias puede ser comprobada por medio de estas secciones de información con vínculo. De esta manera, una materia dada puede ser clasificada de acuerdo con distintos esquemas de clasificación que actúan simultáneamente. De esta manera, Nishikawa expone las iniciativas del consorcio OASIS acerca de la declaración de indicadores de materia publicados.

Mason, uno de los editores de la primera norma ISO 13250, y Vinh Li presentaron en el congreso de ExtremeMarkup en 2002, una comunicación con un título muy sugerente acerca de los orígenes y finalidades de los topic maps: la gestión de clasificaciones de guías y manuales⁷⁰ a partir de una red de materias unívoca que evite el duplicado de las mismas y la ausencia de materias indispensables para el dominio propio de estas guías y manuales.

⁶⁸ BIEZUNSKI, M. The gap between structured and unstructures needs a bridge. En *Extreme Markup Conference*, 2002. En: <http://www.coolheads.com/MBPUBS/extreme02/index.htm> Consultado el 15/12/2009.

⁶⁹ NISHIKAWA, M. Organizing Information in a Corporate Intranet: A Use Case for Published and Internal-Use Subjects in Topic Maps. En *Proceedings of Extreme Markup Languages*, 2002. En: <http://conferences.idealliance.org/extreme/html/2002/Nishikawa01/EML2002Nishikawa01.html>. Consultado el 15/12/2009.

⁷⁰ LÍ, V. ; MASON, D. Topic maps for managing classification guidance. En *Proceedings of Extreme Markup Languages*, 2002. En: www.mulberrytech.com/Extreme/Proceedings/html/2002/Mason01/EML2002Mason01-toc.html. Consultado el 15/05/2009.

Para de Graauw⁷¹ los topic maps proporcionan una manera flexible y coherente de añadir datos aleatorios a las bases de datos. Además los topic maps cuentan con una sintaxis de intercambio predefinido, XTM, para permitir que los datos puedan ser exportados a XML.

Degler y Battle⁷² emplearon topic maps para el diseño de aplicaciones centradas en el usuario para la administración de la seguridad social estadounidense. Con ello se dieron cuenta de que era posible trabajar simultáneamente con varias aplicaciones de manera flexible al generar un contexto indicativo de la aplicación a utilizar de forma pertinente en cada momento.

Garshol presentó un trabajo en la revista *Journal of Documentation*⁷³ en donde contextualiza el papel de los topic maps en los lenguajes documentales. Defiende que los topic maps van más allá de las soluciones tradicionales para la indización y recuperación de información, ya que las propiedades del documento digital facilitan el trabajo con conocimiento más que con datos (ámbitos de las bases de datos relacionales) o información (ámbito de las bases de datos documentales). Esto es así porque el potencial de las relaciones o hipervínculos entre conceptos proporciona un marco contextual que dota a la información de semántica.

Pepper, desde su empresa Ontopia, ha realizado numerosos portales para organismos y empresas de todo tipo en Noruega. Es destacada su labor tanto en la aplicación de topic maps a bibliotecas digitales como portales. Ya en el Congreso de XML Europa de 2004 presentó uno de los primeros trabajos sobre el estado de la cuestión en portales del sector público⁷⁴. Presenta varios ejemplos de cómo se ha aplicado el TAO de los toicmaps, así como editores concretos de topic maps tales como el Omnigator o lenguajes como Zope para topic maps (ZTM).

⁷¹GRAAUW, Marc de. *Using topic maps to extend relational databases*. En: www.xml.com/pub/a/2003/03/05/tmrdb.html. Consultado el 15/12/2009.

⁷²DEGLER, D. ; BATTLE, L. Can topic maps describe context for enterprise-wide applications? En *Extreme Markup Conference*, 2003. En: www.mulberrytech.com/Extreme/Proceedings/html/2003/Degler01/EML2003Degler01.html. Consultado el 15/12/2009.

⁷³GARSHOL, L.M. Metadata? Thesauri? Taxonomies? Topic maps!: Making sense of it all. *Journal of Information Science*, vol. 30, n°4, 2004. En: www.ontopia.net/topic_maps/materials/tm-vs-thesauri.html. Consultado el 15/12/2009.

⁷⁴PEPPER, S. *Towards Seamless Knowledge. Integrating Public Sector Portals in Norway*. En: www.ontopia.net/topic_maps/materials/Towards%20Seamless%20Knowledge.ppt. Consultado el 15/12/2009.

Insalaco y Mason⁷⁵ presentaron en este mismo congreso un ejemplo de aplicación de los topic maps en la intranet de una empresa. Consistía en un topic map con el que era posible localizar con prontitud las herramientas necesarias para un producto demandado por los empleados de la empresa. Para ello, en el topic map se trata con detalle las características de los productos, las series de componentes así como las herramientas disponibles, las habilidades y conocimientos de los distintos empleados para operar con cada producto. Todo ello con el objeto de crear un sistema de gestión del conocimiento exhaustivo y preciso para una empresa que ya empieza a operar con objetos de información digital, pero que necesitan integrarlos en un contexto semántico para poder ser procesados por ordenadores.

Otro trabajo orientado a la recuperación de software y representación de información sobre software con topic maps fue presentado en este mismo congreso por parte de Terrence Brady⁷⁶. Según él se han logrado en su empresa resultados muy satisfactorios empleando topic maps para modelar una colección interdependiente de aplicaciones recolectoras de metadatos.

Lavik y Nordeng⁷⁷ hicieron una de las primeras propuestas aplicativas de toicmaps en educación, que posteriormente daría lugar a un grupo bastante activo en Noruega respecto al tema. Presentan un paquete de herramientas que denominan Brain Bank Learning para el aprendizaje intuitivo de conceptos y contenidos educativos interrelacionados. El núcleo del paquete es BrainBank, la ontología es un topic map construido a partir de un repositorio de documentos educativos. BBL es una aplicación web para el aprendizaje asistido por ordenador que consiste en un editor de topic maps para web que funciona con los navegadores comerciales.

Biezunski en el congreso de Extreme Markup de 2005⁷⁸ sugirió que la denominación de tecnologías semánticas de la información es algo engañosa porque difícilmente una

⁷⁵ INSALACO, T ; MASON, D. Navigating the production maze: The Topic mapped Enterprise. En: <http://shelter.nu/art-007.html>. Consultado el 15/12/2009.

⁷⁶BRADY, T. Representing Software System Information in a Topic map. En *Extreme Markup Conference, 2004*. En: www.mulberrytech.com/Extreme/Proceedings/html/2004/Brady01/EML2004Brady01.html. Consultado el 15/12/2009.

⁷⁷LAVIK, S. ; NORDENG, M. BrainBank Learning. Building topic maps as strategy for learning. En *XML Europe, 2004*. En: www.idealliance.org/papers/dx_xml04/papers/04-02-03/04-02-03.html. Consultado el 15/12/2009.

⁷⁸ BIEZUNSKI, M. *A matter of perspectives: talking about topic maps*. En: www.mulberrytech.com/Extreme/Proceedings/html/2005/Biezunski01/EML2005Biezunski01.html. Consultado el 15/12/2009.

tecnología va a disponer de una cualidad humana como la semántica. Es por ello que más que utilizar el término “ontología” emplea el concepto de “perspectiva”. Con ello quiere enfatizar el hecho de que la información semánticamente integrada depende fundamentalmente del punto de vista adoptado más que de una tecnología. Por ejemplo, los modelos de información tradicional, organizados jerárquicamente, no funcionan adecuadamente cuando se trata de objetos de información digital, haciéndose más adecuado la interrelación de información obtenida de fuentes diversas desde un determinado punto de vista. En topic maps esta perspectiva la otorgan los scopes, concepto que veremos más adelante pero que ejercerá una ventaja clave a los topic maps como modelo para la organización de documentación digital y de gestión del conocimiento con independencia de las características del objeto de información.

1.5.2 Topic maps para servicios web.

Vlist presenta un trabajo en 2001 sobre cómo aplicar los topic maps al diseño de un sitio web⁷⁹. Por entonces estaba cobrando auge la sindicación de contenidos como servicio de valor añadido en portales. La sindicación permite tener al usuario del portal actualizado acerca de las modificaciones que se efectúen en el mismo. Uno de los primeros formatos empleados para la sindicación era el formato RSS 1.0. Vlist partiendo de este formato, añade una taxonomía que organiza los metadatos recogidos en un formato clásico de noticias como XML News y lo lleva a RDF y *topic maps*.

Ahmed presentó un trabajo en el que plantea los topic maps para aplicaciones P2P⁸⁰ útiles para el intercambio de información en intranets corporativas. Aplicación que surge en el marco de la gestión del conocimiento en las empresas y organizaciones como servicio de valor añadido. Consiste en la creación de repositorios de conocimiento alimentados por todos los usuarios de las mismas para poder compartir sus conocimientos y habilidades en los distintos segmentos de la organización.

⁷⁹ VLIST, van der. *Building a semantic web site*. En: www.xml.com/pub/a/2001/05/02/semanticwebsite.htm. Consultado el 15/12/2009.

⁸⁰ AHMED, K. TMSHare. Topic map Fragment Exchange in a Peer to Peer application. En *Proceedings of XML Europe 2003*. En: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.922&rep=rep1&type=pdf>. Consultado el 15/12/2009.

Barta, en esta misma línea presentó posteriormente un protocolo de interacción con topic maps en el congreso de Extreme Markup 2005⁸¹. En esta comunicación presenta un protocolo basado en topic maps siguiendo en el paradigma REST. Esto permite a los usuarios del sitio web acceder a los servicios de un topic map a través de una red con independencia de su plataforma y tecnología de almacenamiento. Para la localización precisa de información adopta características de los lenguajes de consulta topic map, TMQL, pero además apoya la mayor parte de la transferencia de todos los mapas. En contraste a aproximaciones similares también soporta operaciones de actualización.

Bandholtz en “A Taxi in Knowledge Land: A Use Case that Combines Topic maps and Web Services in a Public Portal⁸²” expone un proyecto acerca de la aplicación de un topic map a un servicio web. El empleo principal del topic map es para un portal público, la red de información medioambiental de Alemania. Sin embargo este servicio se proporciona para un uso general en el ámbito de la protección medioambiental. El topic map contiene un tesoro y un nomenclátor. El informe nos ofrece una visión general acerca de la arquitectura, métodos, herramientas empleadas y cuestiones de normalización.

1.5.3 Esquemas, reglas y plantillas.

Grønmo presentó uno de los primeros trabajos acerca de la materia en el congreso de XML Europa de 2000⁸³. En su trabajo discutía la necesidad de un mecanismo para la definición de reglas de edición de topic maps que aseguren la coherencia semántica respecto a los fines para los que fue creado.

Rath y Pepper fueron los primeros en formular los principios y elementos de los topic maps sin un lenguaje técnico de programación y marcado⁸⁴. En una comunicación presentada en el congreso de Markup Technologies de 1999 hicieron una introducción en la que no sólo exponían el significado de cada elemento de los topic maps sino que

⁸¹ BARTA, R. *TMIP, a RESTful Topic maps Interaction Protocol*. En:

www.mulberrytech.com/Extreme/Proceedings/html/2005/Barta01/EML2005Barta01.html. Consultado el 15/12/2009.

⁸² BANDHOLTZ, T. *A Taxi in Knowledge Land: A Use Case that Combines Topic maps and Web Services in a Public Portal*. En *Proceedings of Extreme Markup Languages 2002*. En: <http://www.bandholtz.info/publications/2002-XMLE-TB.pdf>. Consultado el 15/12/2009.

⁸³ GRØNMO, G. *Creating semantically valid topic maps*. En: www.gca.org/papers/xml europe2000/papers/s29-02.html. Consultado el 15/12/2009.

⁸⁴ RATH, H. ; PEPPER, S. *Topic maps: Introduction and Allegro*. En: www.topic-maps.com/content/resources/mt99/hhr-stp.pdf. Consultado el 15/12/2009.

también explicaban conceptos extendidos como “plantillas topic maps”, “clases de jerarquías” y “tipos y reglas para las relaciones de asociación”.

Posteriormente, Rath, en el congreso de XML Europa del año 2000⁸⁵, expuso una serie de características técnicas de los topic maps más relacionado con las aplicaciones reales. Concretamente es interesante por las explicaciones que ofrece de las plantillas topic map, concepto acuñado por el comité de normalización para un concepto necesario pero por entonces no normalizado en la formulación básica del topic map. Además, describe potenciales como, las redes semánticas, tipos de relaciones, propiedades de una relación, las jerarquías de clases y las reglas de consistencia.

Los profesores de la Universidad de Minho Ramalho y Henriques presentaron un sencillo lenguaje para la especificación formal de reglas en la edición de topic maps⁸⁶. Este lenguaje permite expresar las condiciones contextuales en clases de topics y su correspondiente modo de procesamiento. Con XTche, un diseñador de topic maps define una colección de reglas que permite verificar si un topic map concreto es semánticamente válido. Como la evaluación manual de extensos topic maps es imposible, se hace necesario el empleo de un validador automático como éste.

Existen muchas maneras de que los topic maps puedan ser empleados para crear y mantener sitios web comerciales. Ogievetsky⁸⁷ demostró que las transformaciones XSLT pueden ser empleadas para generar páginas HTML con enlaces formulados con topic maps. Los constructos de los topic maps pueden así jugar roles específicos en el proceso de creación automática de HTML.

Passin en el congreso de Extreme Markup 2003 presentó un novedoso trabajo para la aplicación de topic maps a la gestión de bookmarks⁸⁸. Para ello partía de las características anidadas del bookmark, una colección de carpetas que contienen una selección de

⁸⁵ RATH, H. *Making Topic maps more colourful*. En: www.gca.org/papers/xml europe2000/papers/s29-01.html. Consultado el 15/12/2009.

⁸⁶ LIBRELOTTO, G. ; RAMALHO, E. ; HENRIQUES, P. XTche. En: <http://repositorium.sdum.uminho.pt/retrieve/1358/XML2004-XTche.pdf>. Consultado el 15/12/2009.

⁸⁷ OGIEVETSKY, N. Building dynamic web sites with topic maps and XSLT. En *Extreme Markup Conference 2000*. En: <http://ep.open.ac.uk/PubSys/resources/html/ogie1026.html>. Consultado 15/12/2009.

⁸⁸ PASSIN, T. Browser Bookmark Management with Topic maps. En *Extreme Markup Conference 2003*. En: www.mulberrytech.com/Extreme/Proceedings/html/2003/Passin01/EML2003Passin01.html . Consultado el 15/12/2009.

hiperenlaces con sus correspondientes descripciones y agrupados de acuerdo con un tema. Razón por la cual concebía una carpeta como un topic y los hiperenlaces que contiene como topicrole.

1.5.4 Representación de información con topic maps.

El modelo de referencia⁸⁹ para la norma ISO 13250 de Topic maps (RM4TM) proporciona un marco para la definición de aplicaciones topic map que facilita el objetivo de univocidad en la localización de materias no sólo cuando los topic maps que participan en una aplicación para topic maps son aglutinados sino también los topic maps editados con distintas aplicaciones y basados en distintas ontologías son aglutinados.

Otro modelo de representación es TMAPI. Se trata de una propuesta de interfaz de programación común para procesadores de topic maps. El objetivo de TMAPI es permitir a los desarrolladores aprender y usar un solo programa API para trabajar con cualquier motor de procesamiento topic map, mejorando la portabilidad del código y reduciendo la curva de aprendizaje. TMAPI es un software de dominio público y gratuito sin ningún tipo de restricción sobre su uso. Algunos motores que soportan TMAPI son: TM4J, tinyTM, XTM4XMLDB y TMAPI/PHP.

1.5.5 Trabajos sobre la interoperabilidad de Topic maps con metadatos.

En general cuando se habla de interoperabilidad entre topic maps y metadatos cabe hablar de dos ámbitos diferentes de trabajo. Uno es acerca de la interoperabilidad de los topic maps con un esquema para la definición de metadatos como es RDF. Es el campo más antiguo y el más prolífico en cuanto a trabajos sobre la materia debido a la complementariedad que existe entre los topic maps y RDF. Mientras los topic maps se plantean la organización de la documentación a partir de una estructura de materias que el propio usuario puede adaptar a sus demandas de conocimiento más que a las demandas informativas; es decir el usuario ya no tiene demandas informativas sino demandas de conocimiento que satisface a partir de una red de materias enlazadas a recursos de información. Por el contrario el esquemas para la definición de metadatos RDF lo que hace es extender las habilidades del bibliotecario para el análisis formal de los documentos digitales. A partir de un esquema de metadatos en el que se declaran un conjunto de

⁸⁹ RM4TM. En: www.isotopic-maps.org/rm4tm. Consultado el 15/12/2009.

propiedades de los documentos, tan sólo hay que asignar unos valores a cada uno de estas propiedades para efectuar el análisis formal del documento.

Uno de los primeros trabajos acerca de la interoperabilidad entre topic maps y esquemas de metadatos es el trabajo de Pepper⁹⁰ en el que proporciona algunas condiciones que deben darse para que pueda establecerse un marco de interoperabilidad.

En el sitio web del consorcio existe un apartado a las actividades del grupo de trabajo dirigido por Pepper, para la definición de un marco de interoperabilidad entre el modelo RDF y Topic map⁹¹. Entre todos los trabajos destaca un documento acerca de todas las propuestas de interoperabilidad que hasta la fecha se han llevado a cabo.

Uno de los miembros de este grupo, Garshol, tiene publicado un artículo en el sitio web de Ontopia acerca de su visión personal de la interoperabilidad no sólo entre la familia de normas RDF y Topic map, sino de éstas con los distintos lenguajes para el diseño de ontologías⁹². Garshol tiene en la misma página un trabajo anterior acerca del mismo problema de la interoperabilidad RDF vs TM⁹³. Aunque es el precedente, y algo más sintético, del artículo comentado. Es uno de los primeros trabajos donde se efectúa una introducción a cada una de estas tecnologías, destacando las similitudes y diferencias entre ellas. Plantea la posibilidad de crear un marco de interoperabilidad común con los primeros lenguajes de definición de ontologías y los topic maps por su estrecha similitud estructural.

No obstante entre los primeros trabajos presentados en un congreso sobre la materia contamos con la comunicación de Freese⁹⁴, publicada en las actas del propio congreso de markup languages, donde intenta evitar la disputa de intereses entre los defensores de RDF y Topic maps haciendo una comparativa técnica entre ambos modelos.

⁹⁰ PEPPER, S. Navigating haystacks and discovering needles. Introducing the new topic map standard. *Markup Languages: Theory & Practice 1.4*, 1999, p.41-68. En: http://www.ontopia.net/topic_maps/materials/mlangart.pdf Consultado el 15/12/2009.

⁹¹ RDF/Topic maps Interoperability Task Force. En: <http://www.w3.org/2001/sw/BestPractices/RDFTM>. Consultado el 15/12/2009

⁹² GARSHOL, L. *Living with Topic maps and RDF*. En: http://www.ontopia.net/topic_maps/materials/tmrdf.html. Consultado el 15/12/2009.

⁹³ GARSHOL, L. *Topic maps, RDF, DAML, OIL*. En: http://www.ontopia.net/topic_maps/materials/tmrdfoidaml.html. Consultado el 15/12/2009.

⁹⁴ FREESE, E. *Topic Maps vs RDF*. En *Extreme Markup Language 2000*. En: http://www.gca.org/attend/2000_conferences/Extreme_2000. Consultado el 15/12/2009.

Garshol⁹⁵ plantea para la superación de esta disputa un modelo formal para topic maps que denomina Q y que es estructuralmente similar a las representaciones de RDF.

La segunda línea de investigación de la interoperabilidad entre metadatos y topic maps es la promovida recientemente por ISO, donde se formó un grupo para establecer un marco de interoperabilidad entre el esquema de metadatos Dublin Core y topic maps liderado por el profesor Oh.

Moore, en el Congreso de XML Europa 2001, presentó una primera metodología para que RDF pudiera ser empleado en el modelo topic map y viceversa, así como algunas propuestas de modificación para la sintaxis XTM para posibilitar la interoperabilidad entre ambas normas⁹⁶.

En ese mismo congreso, Decker y Lacher realizaron un seminario sobre la web semántica. Defendían, que a fin de prevenir la segmentación de la web en colecciones de recursos incompatibles, la necesidad de compatibilizar RDF con topic maps por medio de la representación de los elementos que constituyen un topic map en RDF. Para ello se mapea el modelo de grafos topic map al modelo de grafos RDF. De esta manera se posibilita la consulta desde aplicaciones propias de RDF⁹⁷.

La Universidad de Bolonia, en el departamento de ciencias de la computación también se lleva tiempo trabajando en una metodología para la interoperabilidad entre los esquemas RDF y topic maps. El equipo del profesor Vitali presentó en el congreso de Extreme Markup de 2003 el trabajo: “Metadata on the web: on the integration of RDF and Topic maps⁹⁸”. En este artículo se presenta una aproximación a la conversión bidireccional entre RDF y los topic maps. Se muestran cómo emplear schemas y la asimilación entre PSI en topic maps y los predicados de RDF como método de integración de ambos esquemas.

⁹⁵GARSHOL, L.M.: Q: A model for Topic maps-Unifying RDF and Topic maps. En *Extreme Markup Conference 2005*. En: www.mulberrytech.com/Extreme/Proceedings/html/2005/Garshol01/EML2005Garshol01.html. Consultado el 15/12/2009.

⁹⁶MOORE, G. *RDF and Topic maps: an exercise in convergence*. En: www.topic-maps.com/topic_mapsrdf.pdf. Consultado el 14/12/2009.

⁹⁷DECKER, S ; LACHER, M. *On the Integration of Topic Maps and RDF Data*. En: www.semanticweb.org/SWWS/program/full/paper53.pdf. Consultado el 14/12/2009.

⁹⁸CIANCARINI, P. ; GENTILUCCI, R. ; PIRRUCCIO, M. ; PRESUTTI, V. ; VITALI, F. Metadata on the web: On the integration of RDF and Topic Maps. En *Proceedings of the Extreme Markup Languages 2003*, Montreal, Canada. En: <http://conferences.idealliance.org/extreme/html/2003/Presutti01/EML2003Presutti01.html>. Consultado el 14/12/2009.

Esta integración es además instrumental para la creación de sencillas herramientas de edición y navegación que puedan ser utilizables por colecciones de metadatos expresadas en ambas lenguas.

Algermissen es uno de los primeros autores preocupados con la interoperabilidad del esquema topic map con los esquemas de metadatos, y más concretamente con Dublin Core. En su artículo⁹⁹: “A processing Model for Dublin Core with HTML”, propone definir un modelo de procesamiento combinado de HTML y el esquema de metadatos Dublin Core. La semántica de Dublin Core la expone por medio de una colección de PSI en un topic map de referencia. En un artículo posterior: “Cooking for the Semantic Web: OWL and Topic map pudding”¹⁰⁰ no pretende proporcionar una formalización completa de topic maps sino que recogiendo los elementos esenciales que constituyen un topic map los relaciona con la sintaxis OWL para la definición de ontologías. Esto sirve de base para posteriores trabajos acerca de la declaración de ontologías a partir de topic maps y que darán lugar al diseño de plugins para la edición de topic maps a partir de editores de ontologías como es el caso de Protégé.

Este trabajo vino a corroborarlo Bernard Vatant, quien en 2004 publicó un trabajo sobre métodos de edición de topic maps con ontologías. En su artículo: “Ontology-driven topic maps”¹⁰¹ defendía la idea de que los topic maps ganarían en efectividad e interoperabilidad a través de esquemas ontológicos específicamente elaborados para el control de topic maps, o bien a través de la declaración de reglas para esquemas de ontologías predefinidas, aunque no expresamente establecidas para ese uso. En cualquier caso, el empleo de ontologías declaradas con sintaxis OWL la consideraba como la opción más plausible para el diseño de topic maps a partir de esquemas de ontologías.

Finalmente Anne Cregan, en el congreso de Extreme Markup de 2005, presentó un modelo de datos para topic maps (TMDM) como una ontología en OWL¹⁰². Esta ontología TMDM es una construcción con elementos de topic maps pero como ontología y en sintaxis

⁹⁹ ALGERMISSEN, J. *Gooseworks.org*. En: www.gooseworks.org/index.html. Consultado el 14/12/2009.

¹⁰⁰ VATANT, B. OWL and Topic map Pudding. En *Mondeca White Paper*, 2003. En: www.mondeca.com/owl/owltm.html. Consultado el 14/12/2009.

¹⁰¹ VATANT, B. Ontology-Driven Topic maps. En *Proceedings of XML Europe, 2004*. En: www.idealliance.org/europe/04/call/xmlpapers/03-03-03.91/03-03-03.html Consultado el 15/12/2009.

¹⁰² CREGAN, A. *Building Topic maps in OWL-DL*. En: www.mulberrytech.com/Extreme/Proceedings/html/2005/Cregan01/EML2005Cregan01.html. Consultado el 15/12/2009.

específica de ontologías. Haciendo así posible que cualquier usuario que desee editar topic maps pueda hacerlo para que también sean interoperables con un programa editor de ontologías.

1.5.6 Diseño de sistemas.

Moore fue de los primeros autores en concebir los topic maps como una tecnología¹⁰³. En el congreso de XML Europa de 2000 presentó una comunicación en la que se centraba más en los aspectos tecnológicos de los topic maps que en la identificación de los elementos conductores del topic map tales como la edición de topics o la fusión de topic maps. Se preocupa en analizar tecnologías como el diseño de software orientado a objetos, diseño de interfaces o búsqueda y mantenimiento de topic maps, más adecuadas a las características intrínsecas de los documentos topic map.

En este mismo congreso Mugnaini¹⁰⁴ coincidió con Moore en destacar el aspecto tecnológico de los topic maps y expuso la posibilidad de un almacenamiento permanente de topic maps por medio del empleo de bases de datos relacionales. Para ello presenta un esquema para una base de datos capaz de almacenar topic maps.

Respecto a las Published Subject Indicator (PSI), se trata de materias publicadas, con sus respectivas urls, que identifican unívocamente aquellos recursos informativos que mejor describen una materia. Una lista de las mismas la podemos encontrar en el portal Easy Topic maps¹⁰⁵.

Una buena introducción sobre la temática la aporta Vatan¹⁰⁶, quien se pregunta por los mejores modos de empleo de PSIs. Vatan nos recuerda que la fusión de topics es un proceso dirigido por los editores de topic maps. Por medio del empleo de los PSIs como elementos vinculantes es posible la construcción de redes semánticas colaborativas de topic maps independientes como alternativa a la unión de topic maps, lo que generaría topic

¹⁰³ MOORE, G. Topic map technology. State of the art. En *XML Europe, 2000*. En: www.gca.org/papers/xml europe2000/papers/s22-04.html. Consultado el 15/12/2009.

¹⁰⁴ MUGNAINI, L. Mapping Topic maps on relational databases. En *XML Europe, 2000*. En: www.gca.org/papers/xml europe2000/papers/s22-04.html. Consultado el 15/12/2009.

¹⁰⁵ *Easy Topic maps*. En: <http://easytopic maps.com/index.php>. Consultado el 15/12/2009.

¹⁰⁶ VATANT, B. Binding points for Subject Identity: The case for standard Published Subject Indicators. En *Extreme Markup 2001*. En: www.mulberrytech.com/Extreme/Proceedings/html/2001/Vatant01/EML2001Vatant01-toc.html. Consultado el 15/12/2009.

maps muy extensos y difíciles de manejar. Este proceso distribuido conduce a una nueva visión de la identidad de una materia sobre la base del conocimiento organizado dinámicamente entorno a estos puntos vinculantes que actúan como el hilo para la encuadernación de un libro.

Otra buena introducción la proporciona Barta¹⁰⁷, de la Bond University en Australia, quien defendía que, de acuerdo con el paradigma de los topic maps, el problema para relacionar un topic con su materia se resuelve empleando direcciones url de materias e indicadores de materia¹⁰⁸ como único modo de relación entre nuestra representación de la realidad con un topic map y los objetos físicos y abstractos a los que se refiere.

Posteriormente los profesores Maicher y Witschel, de la Universidad de Leipzig, presentaron una metodología cuantitativa¹⁰⁹, que denominaron SIM, basada en la asignación de valores a la identidad de una materia. Esta metodología concede más flexibilidad a la representación de materias al dejar de ser descritos los topics a partir de un vocabulario controlado y centralizado de PSIs.

Finalmente, Ahmed¹¹⁰ procuró superar el factor limitador para los *topic maps* de los lenguajes controlados PSI. En el desarrollo de software los modelos de diseño son descripciones estructuradas de las rutinas comunes con las cuales los programadores resuelven sus problemas. Para los editores de topic maps existe un paralelismo. A menudo en una aplicación topic map existe más de una solución para resolver un problema ya que la estructura de los topic maps es lo suficiente flexible para que convivan dos soluciones diferentes. El desarrollo de modelos formales para el diseño de topic maps beneficia a los diseñadores del mismo modo que los diseñadores de software se benefician de los modelos de diseño de software.

¹⁰⁷ BARTA, R. *Is he the one? Subject Identification in Topic maps*. En: http://topic_maps.it.bond.edu.au/docs/21. Consultado el 15/12/2009.

¹⁰⁸ Como veremos un indicador de materia es un documento o recurso informativo seleccionado como representativo de una materia.

¹⁰⁹ MAITCHER, L. ; WITSCHHEL, H. F. *Merging of distributed Topic maps based on the Subject Identity Measure (SIM)*. En: www.lutzmaicher.de/PUBLIKATIONEN. Consultado el 15/12/2009.

¹¹⁰ AHMED, K. *Beyond PSIs. Topic map design patterns*. En *Extreme Markup, 2003*. En: www.mulberrytech.com/Extreme/Proceedings/html/2003/Ahmed01/EML2003Ahmed01.html. Consultado el 15/12/2009.

1.5.7 Sitios web de información especializada sobre topic maps.

topic maps.com¹¹¹.

Este dominio pertenece actualmente a la organización internacional para la normalización ISO y su objetivo es recoger en él todos los PSIs e informaciones acerca de topic maps.

topic maps.org¹¹².

Topic maps.org es un consorcio constituido por organismos, tanto públicos como privados, interesados en el desarrollo de la aplicabilidad del paradigma de los topic maps en la Web a través de toda la familia de especificaciones XML que se puedan necesitar los topic maps.

topic maps.net¹¹³.

Este sitio web informativo acerca de los topic maps, matenido por dos de los principales expertos en topic maps, Biezunski y Newcomb, es la url a la página principal de la empresa InfoLoom sobre sistemas de organización del conocimiento de ambos autores. Es interesante destacar la aplicación Taxmap acerca de un topic map como asistente a los empleados de la oficina de contribución estadounidense.

tmapi.org¹¹⁴

TMAPI (Interfaz para la programación de aplicaciones comunes topic map) es un interfaz de programación para el acceso y manipulación de datos manejados desde un topic map. La especificación TMAPI define una colección de interfaces básicos, comunes e interoperables para que todos los desarrolladores de topic maps puedan procesar los topic maps contruidos con las distintas herramientas de edición de topic maps. Es muy utilizada por desarrolladores de topic maps como Techquila, Ontopia y Empolis.

Bond University¹¹⁵.

¹¹¹ TOPICMAPS.COM. En: www.topicmaps.com . Consultado el 15/12/2009.

¹¹² TOPICMAPS.ORG. En: www.topicmaps.org . Consultado el 15/12/2009.

¹¹³ TOPICMAPS.NET. En: www.topicmaps.com . Consultado el 15/12/2009.

¹¹⁴ TMAPI.ORG. En: <http://tmapi.org> . Consultado el 15/12/2009.

Se trata del sitio web de la Universidad de Bond acerca de topic maps. Contiene numerosos tutoriales, ejemplos de topic maps y enlaces a distintos recursos sobre los mismos. Destaca el trabajo del profesor Robert Barta quien definió la sintaxis para el diseño de topic maps Astma.

Easy Topic maps¹¹⁶.

Se trata de un topic map acerca de topic maps que incorpora wikis sobre los mismos, de forma que cualquier usuario puede añadir información de actualidad o modificar la ya existente.

TMWiki¹¹⁷.

Se trata de un wiki mantenido por Hendrik Thomas, quien define el TMWiki como una wiki acerca de topic maps con el objetivo de ser un punto de acceso al campo de los topic maps. En él se recoge y presenta toda la información que el usuario necesita para la comprensión de los topic maps y de cómo pueden ser utilizados. Además presenta la capacidad de permitir la consulta de topic maps ya contruidos.

1.5.8 Blogs.

Larsblog¹¹⁸

Se trata del blog de Lars Garshol, uno de los desarrolladores más activos de topic maps, quién empezó trabajando sobre el tema con Steve Pepper. Es un blog fundamental para mantenerse actualizado tanto sobre los últimos congresos y seminarios acontecidos sobre la materia como acerca de las últimas normas aprobadas en los comités de normalización, dado que Lars es miembro del subcomité ISO para el desarrollo de topic maps.

Electric Forest¹¹⁹.

¹¹⁵ BOND UNIVERSITY. En: <http://topic-maps.bond.edu.au>. Consultado el 15/12/2009.

¹¹⁶ EASY TOPIC MAPS. En: <http://easytopic-maps.com>. Consultado el 15/12/2009.

¹¹⁷ TMWIKI. En: www.topic-maps.org Consultado el 15/12/2009.

¹¹⁸ LARSBLOG. En: <http://www.garshol.priv.no/blog>. Consultado el 15/12/2009.

¹¹⁹ BLOG ELECTRICFOREST. En: <http://www.altheim.com/ef>. Consultado el 15/12/2009.

Murray Altheim de la Open University tiene este blog acerca de libros, bibliotecas digitales y distintas aplicaciones acerca de la materia. En este blog intervienen numerosos desarrolladores.

Topic map Thoughts¹²⁰.

Blog de Dimitry Bogachev sobre topic maps y especialmente acerca de propuestas y experiencias acerca de TMCL y esquemas para topic maps.

universimmedia¹²¹.

Cómo alcanzar una identificación de materias en una red distribuida.

Techquila Shots¹²².

Opiniones, noticias y enlaces sobre XML, web semántica y gestión de información.

1.5.9 Aplicaciones en el mundo real de los topic maps¹²³.

Real World Applications of Topic maps¹²⁴.

En este artículo se presentan algunas de las primeras aplicaciones al mundo real del modelo de los topic maps en el contexto de una empresa. Aquí se discute las posibilidades de la aplicación y las dificultades encontradas para el diseño y su aplicación.

LmTM¹²⁵.

Se trata de una aplicación para el apredizaje por e-learning de física, ciencias de la computación y matemáticas para la educación superior. La navegación se basa en topic maps contruídos bajo la sinaxis XTM.

¹²⁰ TOPIC MAP THOUGHTS. En: <http://homepage.mac.com/dmitryv/iblog>. Consultado el 15/12/2009.

¹²¹ UNIVERSIMMEDIA. En: <http://universimmedia.blogspot.com>. Consultado el 15/12/2009.

¹²² TECHQUILA SHOTS. En: <http://www.techquila.com/blog>. Consultado el 15/12/2009.

¹²³ Siempre que haya dudas acerca de cualquier ejemplo sobre el empleo de aplicaciones para la construcción de topic maps se ha hecho uso de listas de correo electrónico específicas de topic maps. Esta relación de aplicaciones no pretende de ningún modo ser exhaustiva dado que no hay manera de saber todo lo que los usuarios han escrito hasta ahora acerca del diseño de aplicaciones para la edición de topic maps y sus posibles usos en los muy diversos dominios de aplicación.

¹²⁴ REAL WORLD APPLICATIONS OF TOPIC MAPS. En: <http://www.idealliance.org/papers/xml2001papers/tm> . Consultado el 15/12/2009.

¹²⁵ RITTERSHOFER, A. *Lerner mit Topic maps*. En: <http://www.lmtm.de>. Consultado el 15/12/2009.

Portales noruegos contruídos con topic maps.

- <http://forskning.no> (Consejo de Investigación Noruego)
- <http://forbrukerportalen.no/> (Asociación Noruega de Consumidores)
- <http://matportalen.no/> (Portal del Departamento de Agricultura Noruego)
- <http://itu.no> (Portal del Departamento de Educación sobre tecnologías de la información y comunicación de Noruega)
- <http://www.hoyre.no> (Partido Conservador Noruego)

1.5.10 Tutoriales.

Uno de los mejores tutoriales es el diseñado para los estudiantes de informática de la Bond University (Australia) para que comiencen a diseñar sus propios topic maps. Recoge los conceptos más básicos tanto de los topic maps como del lenguaje de marcado XTM de escritura para topic maps. También es posible escribir el topic map con la sintaxis simplificada LTM (Linear Topic map).

1.5.11 Listas de distribución.

Hemos realizado un uso intensivo de las listas de distribución específicas de los topic maps por tratarse de una fuente muy actualizada de información. El empleo se ha efectuado de dos maneras: a) Cuando teníamos dudas sobre el funcionamiento y aplicaciones de los topic maps hemos enviado consultas a estas listas; b) hemos consultado también las dudas planteadas por otros usuarios de los topic maps y las respuestas que se les han ofrecido.

Las listas utilizadas han sido de dos tipos:

Listas especializadas en las que es posible consultar cuestiones específicas relativas a aplicaciones para edición de topic maps y de normas que lo regulan.

Topic map mailing list. Se trata de una de las primeras listas de distribución acerca de los topic maps, especialmente adecuada para la resolución de dudas relacionadas con aplicaciones para la edición de topic maps. Es interesante su repositorio de consultas realizadas porque a través de él es posible seguir la evolución de los topic maps.

Sc34wg3@isotopic maps.org. Su propósito es facilitar el trabajo técnico por parte de los miembros que componen este subcomité de la organización internacional de normalización encargado de normalizar tanto los elementos que constituyen un topic map como el lenguaje de marcado específico para topic maps. Aunque puede ser consultada por cualquier usuario que lo solicite, sin embargo está restringida la posibilidad de enviar consultas.

TMQL mailing list (tmql-wg). Es una lista de la subcomisión ISO dedicada a la normalización de un lenguaje de marcado específico para consultas en topic maps.

TMCL mailing list (tmcl-wg). Se trata de una lista para la subcomisión ISO dedicada a la normalización de un lenguaje de marcado para la definición de un esquema con el que poder declarar las normas y propiedades de un topic map.

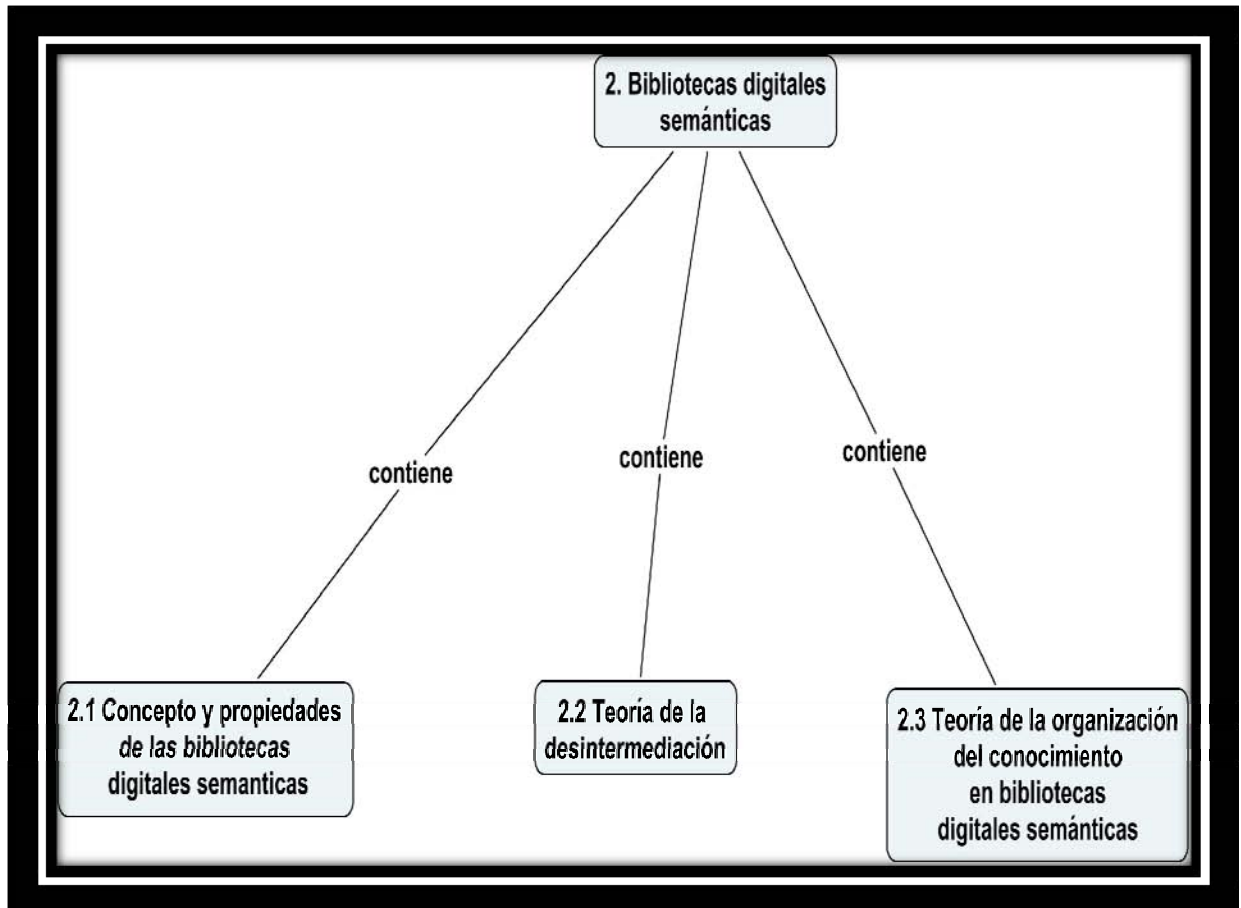
Estas dos últimas listas también permiten el acceso libre a los mensajes por parte de cualquier usuario, pero no así la posibilidad de enviar consultas, ya que éstas son previamente revisadas por el administrador de estas listas.

Listas generales para los usuarios de topic maps en países o regiones geográficas concretas. En España no existe ninguna, pero en Noruega, Alemania, Australia y Nueva Zelanda, donde existen importantes comunidades de usuarios, cuentan con sus respectivas listas de distribución. Sin embargo, la más conocida es la Topic maps Interest Group Australasia (TMIGA) por encontrarse todos los correos en inglés lo que propicia la participación de una comunidad importante de usuarios. Es interesante por los correos acerca del lenguaje simplificado LTM (Linear Topic map) para la declaración de los elementos topic map. Además están expuestas numerosas consultas acerca de la iniciativa para la aplicación de topic maps en el portal del New Zealand Electronic Text Centre (NZETC).

BLOQUE I. Modelo Teórico de los Topic maps en el contexto de la Arquitectura de Información.

CAPÍTULO 2.

Bibliotecas Digitales Semánticas



Introducción.

En este capítulo vamos a introducir el concepto de biblioteca digital semántica como objeto último de la investigación en nuestra tesis y al mismo tiempo tendencia de la investigación en bibliotecas digitales.

Para ello definimos las características diferenciadoras de las bibliotecas digitales semánticas respecto de otros conceptos de biblioteca como las bibliotecas digitales o las bibliotecas electrónicas.

En el segundo epígrafe exponemos un modelo explicativo de esta tendencia evolutiva de las bibliotecas hacia un creciente protagonismo del usuario en la organización y gestión de la biblioteca en el contexto virtual de las mismas.

Finalmente, en un tercer apartado, explicamos el papel de la recuperación por materias en las bibliotecas que en el contexto de la Web Semántica y la Web social alcanzan un papel central por la capacidad que proporcionan a las máquinas para trabajar de forma autónoma en la recuperación de información.

2.1 Concepto y propiedades de las bibliotecas digitales semánticas.

Llegamos al punto final que nos exige el más alerta esfuerzo de atención, porque el tema del libro y del bibliotecario hasta aquí tan manso,..., va a transmutarse de pronto en un drama. Pues bien, ese drama va a constituir la más auténtica misión del bibliotecario. Hasta ahora habíamos topado sólo lo que esta misión ha sido, las figuras de su pretérito. Mas ahora va surgir entre nosotros el perfil de una nueva tarea incomparablemente más alta, más grave, más esencial. Cabría decir que hasta ahora vuestra profesión ha vivido sólo las horas de juego y preludio. Ahora viene lo serio, porque el drama empieza¹²⁶.

Este párrafo extraído del discurso de apertura del 2º Congreso Internacional de la IFLA celebrado en Madrid en el año 1935 y más tarde publicado en la Revista de Occidente, constata la evolución y profesionalización de la tarea bibliotecaria. De hecho su discurso es una breve historia de la labor del bibliotecario con el objeto de evidenciar el oficio de bibliotecario como una misión en la vida.

En efecto, el concepto de biblioteca ha cambiado de forma considerable en el último siglo, fruto de una progresiva profesionalización de sus tareas, favorecido por el impulso que las tecnologías de la información y la comunicación han dado a las Bibliotecas y sus servicios.

Con la aparición de Internet y las posibilidades de digitalización de recursos en los últimos años ha dejado de entenderse la biblioteca como espacio depositario de la colección o como su espacio de consulta. Podemos decir que la biblioteca actual incorpora toda una serie de variedades documentales que están en función de las necesidades de información de sus usuarios; documentos que a través de las tecnologías de la información y la

¹²⁶ ORTEGA y GASSET, J. Misión del bibliotecario. *Revista de Occidente*, año XIII, nº CXLIII, 1935, p.141.

comunicación, están integrando los paradigmas de la biblioteca del presente con la del futuro.

El diseño global de la biblioteca está cambiando; ya que, la biblioteca no sólo recolecta, organiza, provee acceso y preserva información de la forma clásica sino que también trabaja con otros elementos y herramientas como los sistemas electrónicos para la transferencia de archivos (Z39.50), la web o los sistemas marco para la recuperación y localización de información tales como portales, pasarelas o repositorios.

A causa de la inserción de estos nuevos factores en el concepto de biblioteca, así como la adaptación de las TIC a su funcionamiento ordinario, ha surgido un abanico heterogéneo de concepciones sobre el papel de las bibliotecas, generándose una cierta confusión acerca de sus funciones y servicios. Confusión que en este capítulo deseamos resolver, obligados por nuestro objeto de estudio, antes de plantearnos una metodología para la aplicación de un modelo de organización para bibliotecas, Topic maps.

Para contextualizar apropiadamente nuestra concepción de biblioteca vamos a explicar globalmente a cuál nos suscribimos y por qué no utilizamos otras concepciones.

En la literatura de la Ciencia de la Información y Documentación se emplean tres concepciones básicas de biblioteca¹²⁷, pero pensamos que el estado actual de servicios ofrecidos por las bibliotecas nos permiten hablar de cinco concepciones básicas: i) biblioteca electrónica; ii) biblioteca digital; iii) biblioteca virtual; iv) biblioteca híbrida; v) biblioteca digital semántica.

i) La biblioteca electrónica.

Es aquella que cuenta con sistemas de automatización que le permiten la correcta administración de los materiales que integran su colección, principalmente en papel. Asimismo, cuenta con sistemas de telecomunicaciones que le permiten acceder a su información en formato electrónico, de manera remota o local. Principalmente proporciona catálogos que se encuentran en un mismo espacio físico al igual que la colección tanto en papel como en dispositivos magnéticos y electrónicos¹²⁸.

Fundamentalmente incluye servicios automatizados y bases de datos en CD-ROM, así como todo tipo de aparatos electrónicos que permitan el acceso a la información en

¹²⁷ MARTÍNEZ EQUIHUA, S. *Biblioteca digital. Conceptos, recursos y estándares*. Buenos Aires : Alfagrama, 2007, p.14.

¹²⁸ MORALES CAMPOS, E. *La biblioteca del futuro*. México : UNAM, 1996. p. 32.

formato electrónico pero contenida en un soporte físico y no virtual que debe ser organizado y conservado¹²⁹.

ii) La biblioteca digital

Se caracteriza por su tendencia a digitalizar sus colecciones, lo que permite dotarlas de ubicuidad. Suele tratarse de bibliotecas pequeñas en cuanto al número de usuarios, con una colección básica, con un espectro de materias reducido y una cobertura acotada en cuanto a tiempo y temas¹³⁰.

La idea actual de biblioteca digital comprende múltiples definiciones que con el tiempo han ido cambiando. Borgman cree que esta confusión terminológica se debe a la naturaleza interdisciplinar¹³¹ de la investigación en bibliotecas digitales.

Los bibliotecarios ven las bibliotecas como organizaciones que seleccionan, coleccionan, organizan, conservan, preservan, y proporcionan acceso a la información en nombre de una comunidad de usuarios. En este sentido las bibliotecas no están si no empleando otro sistema de distribución para otras formas de documentos. Así el concepto de biblioteca digital connota la idea futura de biblioteca y al mismo tiempo aporta un sentido de continuidad respecto a lo que hasta ahora han sido las bibliotecas.

Así una de las primeras definiciones de biblioteca digital es la aportada en el informe final de una comisión especializada de bibliotecarios entre los cuales se encontraba Borgman para la National Science Foundation. En ella se recoge la definición aportada en el informe acerca de la conceptualización de la biblioteca digital¹³²:

Las bibliotecas digitales son una colección de recursos electrónicos y capacidades tecnológicas asociadas para la creación, búsqueda y empleo de información. En este sentido son una extensión y mejora de los sistemas de almacenamiento y recuperación de información, al manipular datos digitales capaces de procesar información en cualquier medio (texto, imagen, sonido,..) a través de redes distribuidas. El contenido de las bibliotecas digitales incluye datos y metadatos descriptivos de las distintas propiedades de los datos,..., y metadatos que consisten en relaciones con otros datos o metadatos, tanto internos como externos a la biblioteca digital.

¹²⁹ AMAT I NOGUERA, N. *La biblioteca electrónica*. Madrid : Fundación Germán Sánchez Ruipérez, 1990.

¹³⁰ Id.

¹³¹ BORGMAN, C. L. *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*. Cambridge : MIT Press, 2000, p.16

¹³² Ibid. p.17.

Las bibliotecas digitales son construidas, reunidas y organizadas por una comunidad de usuarios, y sus capacidades funcionales proporcionan las necesidades informativas y usos de esa comunidad. En este sentido son una extensión, mejora e integración de una variedad de bibliotecas físicas donde los recursos son seleccionados, recogidos, organizados, preservados y accesibles en apoyo de una comunidad de usuarios. Estas instituciones informativas incluyen todo tipo de bibliotecas, pero las bibliotecas digitales permiten extender sus capacidades a otras muchas comunidades de usuarios”.

Como vemos, se entiende la biblioteca digital como una colección organizada de documentos almacenados en formato digital que ofrece sus servicios de búsqueda y recuperación de información a una comunidad más amplia de usuarios. El medio digital le permite a las bibliotecas librarse del tan ansiado espacio físico que representan los edificios y estantes, y hace posible la personalización de los recursos y servicios¹³³.

Entre las concepciones generales que se han asignado a las bibliotecas digitales se encuentran las siguientes:

- Una biblioteca digital es un conjunto de elementos de cómputo digital, mecanismos de comunicación y programación necesarios para reproducir, emular y extender los servicios provistos por bibliotecas convencionales como son la catalogación, búsqueda, y difusión de la información¹³⁴.
- Uno de los primeros autores en ofrecer una sistematización de las propiedades de las bibliotecas digitales fue Cleveland. Para él, las bibliotecas digitales son la cara digital de las bibliotecas tradicionales que incluyen tanto colecciones en soportes tradicionales como colecciones digitales. Así que abarcan tanto materiales electrónicos como en papel¹³⁵.

Además de incluir materiales digitales que existen fuera de los límites físicos de esa biblioteca, se incluyen todos los procesos y servicios propios de la biblioteca pero no sólo adaptados al formato digital sino también a la nueva realidad de la ubicuidad de los recursos

¹³³ ARMS, W. Y. Key concepts in the architecture of the digital library. *D-Lib Magazine*. En: www.dlib.org/dlib/July95/07arms.html. Consultado el 15/12/2009.

¹³⁴ GLADNEY, H. M. Digital library: Gross structure and Requirements. En *Proceedings of the First Annual Conference on the theory and practice of Digital Libraries*. Blacksburg : Virginia Polytechnic Institute, 1994.

¹³⁵ CLEVELAND, G. Digital libraries: Definitions, issues and challenges. En *IFLANET: Activities & Services*. En: www.ifla.org/VI/5/op/udtop8/udtop8.htm. Consultado el 15/12/2009.

informativos. Por ello la Association of Research Libraries (ARL) afirma que las colecciones de la biblioteca digital no están limitadas a los documentos impresos, sino que se extienden a los documentos digitales, que no pueden ser representados o distribuidos en formato impreso¹³⁶. En este sentido añaden que la biblioteca digital no es una sola entidad sino que requiere de tecnología para integrar diferentes recursos y tiene como meta el acceso universal.

iii) La biblioteca virtual.

Permite el acceso a la colección desde puntos remotos sin necesidad de contar físicamente con la misma. Fundamentalmente, se ha empleado en ámbitos de servicios de información especializada ya que administra el acceso a información remota que se encuentra en colecciones que poseen otros puntos de servicio, pero a la cual se puede acceder mediante las telecomunicaciones¹³⁷.

Un claro ejemplo de ello es la biblioteca virtual¹³⁸ creada por Tim Berners-Lee. Desde la misma es posible acceder no sólo a una colección de recursos en red, que por supuesto se encuentran en formato digital, y de los que se ofrece un comentario o resumen del mismo como antecedente de los metadatos. Además se estructura el acceso a una materia como un lenguaje controlado, proporcionando la definición de la materia de búsqueda como si se tratara de un glosario, así como relaciones jerárquicas, de equivalencia y asociativas siguiendo la estructura de un tesoro.

¹³⁶ *Definition and purpose of a digital library*. En: www.ifla.org/documents/libraries/net/alrdlib.txt. Consultado el 15/12/2009.

¹³⁷ MARTÍNEZ EQUIHUA, S. Op. cit. p. 21.

¹³⁸ *The www virtual library*. En: <http://vlib.org/index.en> Consultado 04/12/2009.

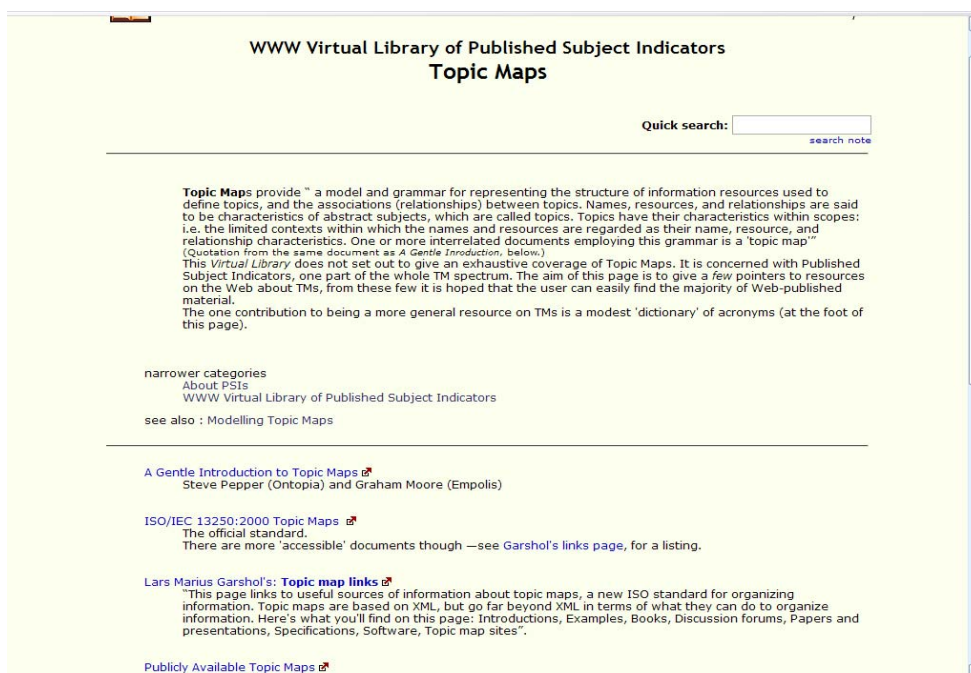


Figura. 4. Un ejemplo de Biblioteca Virtual: The WWW Virtual Library.

Se diferencia de una biblioteca electrónica en que los contenidos de la biblioteca electrónica no siempre tienen que ser virtuales, pueden tratarse de recursos digitales los datos de una base de datos accesibles a través de una intranet o únicamente desde un único ordenador. Realmente se trata de una concepción de la biblioteca proporcionada en dos etapas evolutivas diferentes.

Por otro lado la biblioteca digital pasa no sólo a permitir el acceso a colecciones de documentos digitales sino que permite la interacción del usuario con los mismos a través de la red empleando arquitecturas de información integradoras de un conjunto de servicios de valor añadido como puedan ser un OPAC, bases de datos, revistas electrónicas, etc. que permiten un acceso ubicuo¹³⁹.

iv) El concepto de biblioteca híbrida.

Surge como consecuencia de la lenta adaptación de las características de las bibliotecas físicas a las propiedades de las bibliotecas digitales. Durante mucho tiempo la biblioteca digital era una visión de futuro, pero la realidad era que gran parte de los proyectos

¹³⁹ SAORÍN PÉREZ, T. *Los portales bibliotecarios*. Madrid : Arco Libros, 2004, p. 37

tecnológicos que se deseaban impulsar pretendían potenciar los servicios ya existentes en la biblioteca¹⁴⁰.

El origen del concepto de biblioteca híbrida se basa en un programa de bibliotecas electrónicas, eLib, del Reino Unido. La función de este servicio era unir la biblioteca tradicional con los servicios propios de la biblioteca electrónica, digital o virtual que mejoraban la búsqueda y recuperación de información. Esta idea se basa en la necesidad de proporcionar un acceso simultáneo de los usuarios tanto a los recursos de información electrónicos como a los tradicionales¹⁴¹.

Una biblioteca híbrida es la integración de tecnologías: electrónica, digital, o virtual más los productos impresos y servicios en espacio físico junto a las funciones históricas de ésta. La agencia británica UKOLN para la investigación y provisión de servicios a las redes digitales de bibliotecas, en su programa MODELS¹⁴², extendió el concepto de biblioteca híbrida más allá del dominio de la biblioteca para pasar a hablar de “ambiente de información híbrido”.

Definen un ambiente de información híbrido como un rango apropiado de servicios de información heterogéneos que se presenta al usuario de manera íntegra con un solo interfaz. Su misión es la inclusión tanto de servicios distribuidos locales como remotos, tanto para documentación impresa como digital.

De hecho, un claro ejemplo de los problemas de falta de integración que se procuraban resolver con este programa se encontró en 1997 en la Biblioteca Nacional de Australia. Se identificaron problemas de integración en el ambiente de información que dificultaban el empleo coordinado de los servicios ofrecidos por la biblioteca. Por ello, se diseñó una arquitectura de información teniendo en cuenta un ambiente de información híbrido que permitiera a los sistemas proporcionar un modelo integrado¹⁴³.

Estos problemas se debieron a la disposición de nuevos servicios por parte de la biblioteca digital, que incrementaron la dificultad de integración e interoperabilidad entre los mismos, especialmente entre distintas bibliotecas, o incluso dentro de una misma biblioteca cuando se procuraba la integración de recursos locales y remotos¹⁴⁴.

¹⁴⁰ VILLA BARAJAS, H. ; ALFONSO SÁNCHEZ, I. Biblioteca híbrida: el bibliotecario en medio del tránsito de lo tradicional a lo moderno. *Acimed*, vol. 13, n°2, 2005. En: http://bvs.sld.cu/revistas/aci/vol13_2_05/aci05_05.pdf. Consultado el 15/12/2009.

¹⁴¹ PEARCE, J. ; CATHRO, W. ; BOSTON, T. The challenge of integrated access: the hybrid library system of the future. En *10º VALA Biennial Conference and Exhibition*. Melbourne, 2000. En: www.nla.gov.au/nla/staffpaper/jpearce1.html. Consultado el 15/12/2009.

¹⁴² MODELS (Moving to distributed environments for library services). En: <http://www.ukoln.ac.uk/dlis/models/requirements/overview/>. Consultado el 15/12/2009.

¹⁴³ Id.

¹⁴⁴ Id.

Una de las cuestiones que se plantea el bibliotecario en el ambiente de información híbrido es cómo navegar en él. El empleo de una misma herramienta para documentos en formato impreso y/o digital implica dificultades para proveer un acceso organizado a unos documentos que según el formato en el que se encuentren presentarán unas características u otras; es decir, la motivación subyacente al concepto de biblioteca híbrida es la necesidad de hacer frente a la diversidad, problema relevante si lo que perseguimos es la unificación de sistemas para el aprovechamiento de la ubicuidad en la información digital.

Muchas de las bibliotecas que utilizamos actualmente son híbridas, y es común encontrarlas con servicios tradicionales para una colección en formato impreso junto con servicios de acceso electrónico a un repositorio de documentos y al mismo tiempo servicios de acceso a documentos digitales hospedados en servidores externos a la propia biblioteca. Un claro ejemplo de ello es el sitio web de la biblioteca nacional de Australia¹⁴⁵. Como vemos no sólo es posible acceder al catálogo para la consulta de las referencias de la colección impresa que constituye la colección física de la biblioteca, sino que también podemos acceder a las referencias de documentos en otros formatos como las microformas, los documentos en formato magnético como los audiolibros; o sencillamente acceder por medio de pasarelas a los recursos informativos digitales, bien referencias bien documentos primarios¹⁴⁶, hospedados en servidores ajenos a la propia biblioteca.

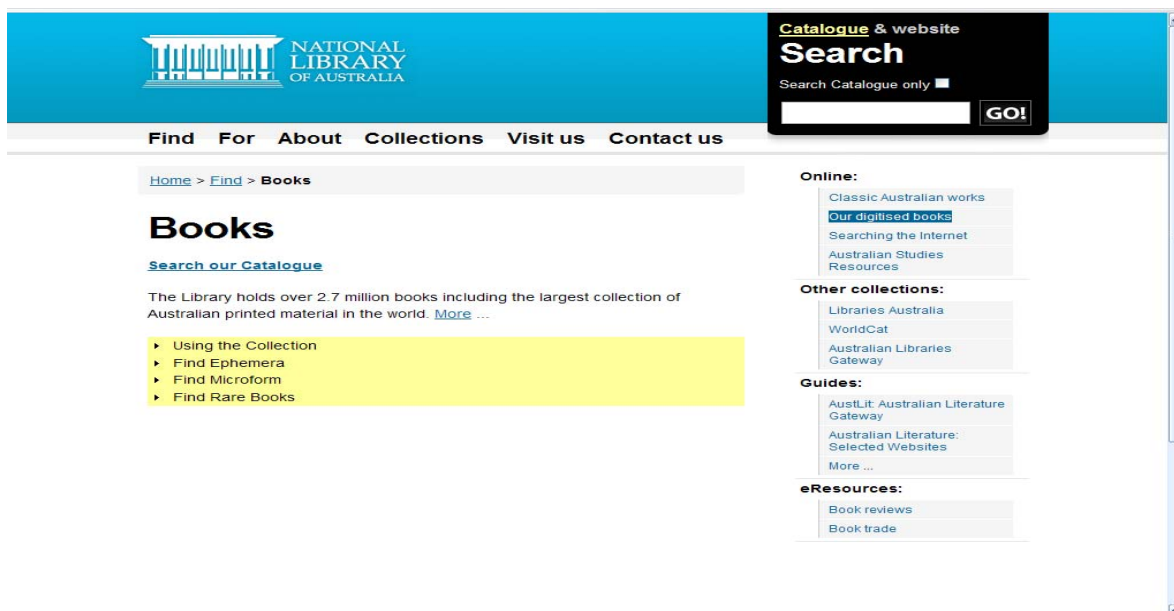


Figura. 5. Biblioteca Nacional de Australia. Ejemplo de Biblioteca Híbrida

¹⁴⁵ NATIONAL LIBRARY OF AUSTRALIA. En: <http://www.nla.gov.au/find/books.html> Consultado el 15/12/2009.

¹⁴⁶ Con ello queremos significar la complejidad de la documentación a la que se permite el acceso dado que ya no se trata de documentación únicamente textual sino que puede tratarse de cualquier tipo de documento como grabados, dibujos, vídeos, etc.

v) Biblioteca digital semántica.

Cuando el espacio documental está dominado por el soporte impreso, la manera de afrontar el espacio es básicamente la misma en todos los casos. No cambia la estructura de la biblioteca sino que con independencia del grado de adaptación tecnológica sigue siendo la misma que ofrecía sus servicios tradicionales. La biblioteca es el agente entre los usuarios y la información, pero con el advenimiento de las nuevas tecnologías, ha comenzado a adoptar nuevas formas adaptadas a la coexistencia de medios impresos y digitales, creándose un nuevo espacio documental que es multilateral por presentarse en ambos formatos.

Las bibliotecas digitales son uno de los recursos más recurridos en Internet para proporcionar un acceso ordenado a una colección o repositorio de contenidos seleccionados con un propósito concreto. A medida que se va produciendo más información y ésta se distribuye a través de la red, más vigencia cobran las tecnologías centradas en el usuario. Y las bibliotecas digitales no son ajenas a esta realidad, su tecnología, su arquitectura, sus normas de diseño deben modificarse para adaptarse a las nuevas necesidades de sus usuarios, con el objetivo último de proporcionarles una mejor experiencia y una aplicación del creciente número de normas que la industria publica a fin de lograr la interoperabilidad entre sistemas.

La próxima generación de bibliotecas digitales combina las soluciones tecnológicas orientadas a las redes sociales, como puede ser P2P, con la capacidad de localización semántica de contenidos.

Las bibliotecas digitales semánticas, más conocidas como las “Library 2.0” siguiendo la tendencia actual de denominar a cualquier evolución tecnológica relacionada con internet 2.0, persiguen fundamentalmente siete objetivos¹⁴⁷:

- i) Cualquier usuario puede hacer uso de ellas,
- ii) es posible acceder a todo documento descrito por una materia incluida por la biblioteca,
- iii) se puede acceder a la biblioteca en cualquier momento,
- iv) es accesible desde cualquier lugar,

¹⁴⁷ AGOSTI, M. [et al.] *Evaluation and Comparison of the service architecture, P2P, and grid approaches for DLs*. Munich : DELOS, 2006. En: www.delos.info. Consultado el 14/12/2009.

- v) se caracteriza por un interfaz de suario multimodal usable,
- vi) proporciona un modo eficiente y efectivo de acceso, y
- vii) emplea múltiples aplicaciones interoperables y dispositivos interconectados.

Blyberg define cinco componentes principales de la Library 2.0¹⁴⁸:

- a) emplea un software abierto,
- b) un software social,
- c) un OPAC integrado,
- d) normas abiertas,
- e) una única clave de entrada.

Los elementos comunes a ambas definiciones son apertura y usabilidad de la Library 2.0 o biblioteca digital semántica.

Proporcionar acceso a la información es la función primaria en toda biblioteca, tanto si es digital como si no lo es. Sin embargo las tecnologías semánticas pueden ofrecer soluciones más eficientes para el acceso a los contenidos. Las etiquetas, descriptores o metadatos proporcionan al usuario no sólo una importante capacidad para elaborar estrategias de búsqueda más elaboradas, sino también la posibilidad de localizar información por símiles semánticos, o el empleo de sistemas de filtrado colaborativo¹⁴⁹.

Por tanto vamos a plantearnos la inserción de tecnologías semánticas a fin de poder mejorar la usabilidad de las bibliotecas digitales mediante el empleo de servicios de asistencia social y semántica al usuario que permitan liberar al bibliotecario de algunas de sus labores de asistencia, dotando al usuario de una mayor autonomía en la biblioteca digital. Sin embargo, cuando se incide especialmente en el empleo de la semántica para la localización de contenidos en una biblioteca digital preferimos hablar de biblioteca digital semántica en vez de Library 2.0.

Las bibliotecas digitales semánticas pueden entenderse desde perspectivas muy diferentes a causa de su naturaleza interdisciplinar, pero en general podemos entenderlas como una combinación de tecnologías de tratamiento de la información y documentos digitales de muy diversa índole. Según esto se necesita un modelo teórico y práctico capaz de integrar

¹⁴⁸ BLYBERG, J. *Library 2.0 websites: Where to begin?* En: www.blyberg.net Consultado el 14/12/2009.

¹⁴⁹ MEGHINI, C. ; SPYRATOS, N. Information Access in digital libraries: Steps towards a conceptual schema. En IOANNIDIS, Y. ; SCHEK, H.J.; WEIKUM, G. *Future Digital Library Management Systems: System Architecture and Information Access*. Munich : DELOS, 2005, pp.48-53. En: http://www.delos.info/files/pdf/Proceedings/Dagstuhl_2903_010405/delos-dagstuhl-handout-all.pdf Consultado el 13/12/2009.

las complejas características que surgen de la interacción de los elementos que participan en una biblioteca digital semántica.

Podemos decir que el modelo arquitectónico de una biblioteca digital semántica deberá resolver los siguientes problemas:

- Integrar información basándose en distintos esquemas de metadatos, vocabularios controlados mediante el empleo de información contextualizada con una red de relaciones que doten de un nivel semántico mínimo a la información
- Proporcionar interoperabilidad con otros sistemas tanto en metadatos como en comunicación. Pensamos que es posible esta interoperabilidad por medio del empleo de RDF y topic maps.
- Proporcionar interfaces de búsqueda y navegación más usables por el concurso de herramientas de inteligencia artificial, perfiles de usuario y contextualización semántica de los contenidos.

Para alcanzar estos objetivos se necesitaría en primer lugar definir un modelo orientado a objetos más amplio¹⁵⁰. El modelo debe hacer uso de la conectividad e hipertextualidad del documento digital. Debe ser capaz de proporcionar una tipología compleja de objetos de información capaz de representar su diversidad. Además el modelo debe ser capaz de proporcionar una variedad de relaciones entre los objetos así como permitir la reutilización de los objetos empleados en otros modelos. Finalmente debe ser capaz de proporcionar al usuario interactividad con los objetos para que éste pueda seleccionar y agrupar los objetos que le resulten más pertinentes a la resolución de sus demandas informativas.

En segundo lugar se necesitaría que incluyera los siguientes servicios básicos:

1. Búsqueda, Navegación y servicios de asistencia a la localización de información. El objetivo clave de la biblioteca digital semántica es el reparto de las tareas de localización de información. Los usuarios deben ser capaces de explotar la información que se extrae por estar interconectada mientras navegan y localizan los recursos¹⁵¹. Los motores de búsqueda deben ser capaces de adaptar las consultas a los perfiles de usuarios que las realizan, explotando las relaciones semánticas entre los resultados. Finalmente la biblioteca digital semántica debe ser capaz de

¹⁵⁰ LAGOZE, C. ; PAYETTE, S. ; SHIN, E. ; WILPER, C. Fedora: an architecture for complex objects and their relationships. *International Journal on Digital Libraries*, vol. 6, nº 2, 2006, pp.124-138.

¹⁵¹ KRUK, S. R. [et al.] *Tutorial on semantic digital libraries at ESWC'2007*. En: <http://library.deri.ie/resource/goljrewV>. Consultado el 15/12/2009.

recomendar distintos servicios basándose bien en el contexto o bien por medio de un trabajo colaborativo de la red social de la biblioteca empleando folksonomías, especialmente si deseamos incorporar servicios propios de la web social como blogs que permiten al usuario ocupar el centro de la organización de la biblioteca¹⁵².

2. Servicios para la adición de recursos informativos. Una de las principales características de las bibliotecas digitales semánticas es la introducción de metaetiquetas que mejoran la calidad de la información proporcionada durante el proceso de introducción de recursos. El objetivo es que este tipo de bibliotecas sean capaces de proporcionar tanto metaetiquetas de forma automática como las incorporadas por el usuario. Las metaetiquetas de los usuarios son la tecnología clave para implicar a los usuarios en el proceso de difusión tanto de los servicios como de la colección de la biblioteca.
3. Servicios de difusión. En este tipo de bibliotecas se ayuda a los usuarios a acceder a los metadatos en distintos formatos a través de máscaras de servicios y contenido. Los servicios de etiquetado extensibles y personalizables entre bibliotecas digitales distribuidas constituyen una parte esencial en su proceso de recuperación de información. Esta integración permite mejorar la experiencia del usuario en relación con el sistema de la biblioteca digital.
4. Servicios proveedores de interoperabilidad, que incluyen la compatibilidad de contenido, metadatos y servicios. En primer lugar debemos matizar que en la literatura científica se identifican varios niveles de interoperabilidad: técnica, informativa y social. La interoperabilidad de servicios puede verse mejorada por un adecuado modelo de gestión de metadatos.
5. Integración de documentación. Las bibliotecas digitales semánticas y los servicios que proporcionan son en muchos casos demasiado complejas para ser fácilmente comprensibles por usuarios no expertos. Se trata de un requisito vital desde la perspectiva de la usabilidad de la biblioteca el contar con una documentación que asista al usuario en la consulta de los servicios ofrecidos por la biblioteca.

Un caso concreto de biblioteca digital semántica es JeromeDL¹⁵³. La investigación inicial acerca de bibliotecas digitales dio lugar al diseño y aplicación de JeromeDL. La

¹⁵² MOREIRO GONZÁLEZ, J. A. La representación y recuperación de los contenidos digitales: De los tesauros conceptuales a las folksonomías. En TRAMULLAS, J. *Tendencias en documentación digital*. Gijón : Trea, 2006. p.104.

¹⁵³ JeromeDL. *E-library with semantics*. En: <http://www.jeromedl.org/> Consultado el 15/12/2009.

investigación de entonces acerca de redes sociales en línea y descubrimiento de información proporcionó nuevas propiedades que fueron aplicadas en JeromeDL.

En esta biblioteca digital se describe cada recurso empleando tres tipos de metadatos: a) estructurales, b) bibliográficos, c) comunitarios. Dispone de servicios que difunden cada uno de estos tipos de información. El etiquetado basado en metadatos y ontologías bibliográficas sirve para ofrecer a los usuarios una representación única de los recursos de información.

La investigación actual de la biblioteca digital JeromeDL se basa en la combinación de ontologías para poder ofrecer al usuario un conjunto de servicios semánticos. Al aplicarse este estudio se averiguó que la gestión de contenidos en bibliotecas digitales puede ayudar al desarrollo de la Web semántica si podemos introducir las características de la misma en su organización, lo que daría lugar a una biblioteca digital semántica.

Se trata de una biblioteca que hace uso tanto de la web semántica como de las tecnologías ligadas a las redes sociales con el objeto de mejorar la interoperabilidad y la usabilidad de la biblioteca digital. Es un proyecto conjunto de la biblioteca universitaria de Gdansk y el laboratorio DERI (Digital Enterprises Research Institute) de Irlanda. Los fines de esta biblioteca eran apoyar el legado de las bibliotecas clásicas en cuanto a conservación de sus colecciones, proporcionar características de navegación más centradas en el usuario, facilitar la búsqueda eficiente de información, reconocer los múltiples formatos en los que se encuentran los documentos digitales, proporcionar seguridad en el acceso a su colección y permitir la comunicación con otros sistemas gestores de bibliotecas digitales.

JeromeDL utiliza un software que hace posible emplear los mismos servicios en plataformas heterogéneas como puedan ser la gestión del perfil de usuario, la búsqueda o la navegación. Así la descripción de los contenidos de los recursos, bien en MARC21 o a través de índices se unifica con una ontología denominada Jerome aplicada a un conjunto de repositorios de información y que, incluso, puede ser aplicada para la recuperación en una red de bibliotecas digitales.



Figura. 6. JeromeDL. Ejemplo de Biblioteca Digital Semántica.

2.2 Teoría de la desintermediación.

Una de las paradojas en la identificación del rol intermediario de las bibliotecas es que tiene lugar con el telón de fondo de la desintermediación. Este término se emplea para designar el proceso por el cual los usuarios son alentados para interactuar directamente con los servicios¹⁵⁴. En bibliotecas, el concepto de autoservicio se comprende como un proceso de desintermediación que elimina la necesidad de personal intermediador entre el sistema y el usuario. Este proceso supone, además de ventajas económicas, beneficios para los usuarios reduciendo las colas, o que puedan acceder a un abanico de posibilidades y servicios adicionales que son difíciles de ofrecer individualmente. Si bien, muchos clientes prefieren tratar con personal bibliotecario antes que la interacción directa con un OPAC.

En este sentido Brophy entiende las bibliotecas como intermediarios desintermediados¹⁵⁵ para referirse a la evolución de las bibliotecas hacia un diseño de sus servicios centrados en la autonomía del usuario, en la capacidad del usuario para acceder en todo momento a cualquier tipo de documento con independencia del espacio físico en que esté hospedado de forma autónoma, a través de un interfaz usable e intuitivo que permita al usuario interactuar de forma autónoma con el sistema.

¹⁵⁴ FOURIE, I. Should we take disintermediation seriously? *The electronic library*, vol. 17, nº1, 1999, pp. 9-16.

¹⁵⁵ BROPHY, P. *The library in the twenty-first century*. London : Facet, 2005, pp. 90-93

Es posible entender la desintermediación desde otro punto de vista siempre y cuando aceptemos que los límites entre los usuarios y el personal bibliotecario se diluyan. En este marco¹⁵⁶:

Los usuarios participan en el servicio, porque pueden alterar el producto que les ofrece y el modo en el cual se les ofrece en respuesta a la situación como se desarrolla. La tecnología de la información permite que los servicios pasen de ser relaciones cliente-empleado a la interacción directa cliente-servicio. Resulta provechoso cuando se analizan estas interacciones cliente-servicio si entendemos al cliente en cierta manera como un empleado gratuito. La comunidad de usuarios disponen así de los servicios por sí mismos. Esto tiene implicaciones para la mejora de la calidad proporcionada por el sistema dada la capacidad para participar en el diseño y alimentación del sistema se añade a la mejora que hagan los propios empleados.

Si bien se está restando protagonismo a la relación entre el bibliotecario y sus usuarios, aquellos siguen valorando su intermediación directa tanto en nuevos servicios como en otros en los que no había una dedicación especial como son los servicios de referencia. Por tanto, en contra de lo que podría suponerse, con la desintermediación los servicios de atención al usuario cobran fuerza, dado que los usuarios que ahora entran en una biblioteca física necesitan más ayuda que antes.

Gorman¹⁵⁷ incide en este punto cuando habla del componente vital persona a persona que ha caracterizado los servicios de referencia a lo largo de la historia. Estamos en una época en la cual los valores humanos son cuestionados y el contacto humano pasa consecuentemente a ser más apreciado a medida que éste se vuelve más escaso¹⁵⁸.

Teniendo esto en cuenta, nos parece pertinente el empleo del término “intermediario” para describir el rol de la biblioteca digital semántica, pero teniendo en cuenta sus limitaciones. Limitaciones en cuanto al funcionamiento de algunos servicios que se organizan en nombre del usuario sin su concurso, especialmente en aquellos centros en los que sus usuarios son desconocedores del funcionamiento adecuado de los mismos.

¹⁵⁶ COULLING, K. *Quality management for information and library managers*. London : Gower, 1996, p. 23

¹⁵⁷ GORMAN, G. E. *The digital factor in library and information services : International Yearbook of Library and Information Management 2002-2003*. London : Facet, 2002. pp. 45-48.

¹⁵⁸ Id.

Al mismo tiempo se ha evidenciado que muchos servicios han sufrido un cambio paralelo. En los entornos en red, los usuarios de los servicios se han visto favorecidos por un principio de servicios distribuidos basado en la difusión de tareas ya previamente creadas, por el cual los usuarios escogen organizar sus tareas por sí mismos. Hasta ahora se difundían servicios por medio de un interfaz propio, pero ahora se tienen que ofrecer para que sean accesibles desde cualquier interfaz como es el caso de los servicios por Z39.50.

La desintermediación se entiende en función del papel jugado por el mediador entre la información y su comunidad de usuarios. Supone la búsqueda de información por parte de un usuario final sin la necesidad de terceras partes. Según Gellman¹⁵⁹:

la desintermediación aplicada a las bibliotecas significa la evolución de la información desde depósitos físicos centralizados hacia fuentes alternativas accesibles directamente desde ordenadores y redes.

Potenciar al usuario final para el acceso a la información supone dotarle de unas habilidades necesarias para la satisfacción de sus demandas informativas. Con esta potenciación deberían ser menos dependientes del bibliotecario, lo que no significa que su papel como mediador de información haya pasado a un segundo plano: no todos los usuarios finales tienen tiempo o interés en realizar sus propias búsquedas. Aunque hay una conexión entre desintermediación y potenciación del usuario final, ésta no implica necesariamente la desintermediación¹⁶⁰.

Con el advenimiento de Internet y el aumento del acceso y conocimiento de la información parece inevitable que los usuarios finales realicen sus propias búsquedas de información. Evidentemente habrá algún tipo de desintermediación, pero su extensión dependerá de factores tales como el contexto social, las políticas organizacionales de búsqueda para usuarios finales, la tecnología disponible, y los servicios ofrecidos por servicios de información concretos; además de las reacciones que los documentalistas y los gestores de los servicios de información ejerzan en el curso de la desintermediación.

Los bibliotecarios siempre han pretendido mejorar el acceso a la información organizándola y proporcionando información a los usuarios conforme a sus necesidades.

¹⁵⁹ GELLMAN, R. Disintermediation and the Internet. *Government Information Quarterly*, vol. 113, nº1, pp.1-8.

¹⁶⁰ BROPHY, P. *The library in the twenty-first century*. London : Facet, 2007, p.78.

Los servicios de información también buscaban ayudar al usuario final a realizar su trabajo con mayor efectividad. El cambio en los requerimientos del usuario final, su acceso a la información y sus necesidades, ha provocado un cambio paralelo en el papel de los mediadores. Esto es importante para los mediadores que tratan de mejorar el acceso social a información de calidad. Si más usuarios finales son capaces de asumir la responsabilidad de sus propias búsquedas de información, la atención de los mediadores puede orientarse hacia otras cuestiones importantes. La potenciación del usuario final puede entonces ayudar a asegurar la continuidad del mediador y promover usuarios finales bien informados¹⁶¹.

En el medio cambiante de todas las disciplinas relacionadas con la información ha habido muy distintas reacciones al cambio de entorno. En el ámbito de las bibliotecas las reacciones están influidas no solo por los individuos sino también por la comunidad de usuarios a la que se dirigen los servicios, el modelo de gestión y la cultura organizacional¹⁶².

Los bibliotecarios pueden reaccionar, tal y como se muestra en la tabla siguiente, de la siguiente manera a la desintermediación¹⁶³:

1) Actitud Pasiva.	2) Actitud: adopción de opciones distintas.	3) Actitud: aplicación de las propias opiniones.	4) Actitud: fijación de escenarios y realización de prospectiva.	5) Actitud: facilitado de respuestas alternativas a escenarios similares.	6) Actitud: reflexión sobre comportamientos pasados.
Ignorancia de los cambios e implicaciones.	Optar por algo completamente diferente.	Especulación sobre el problema.	Realización de escenarios de futuro	Análisis de otras respuestas a escenarios similares	Reflexión acerca de los comportamientos pasados
Asunción incertidumbre futuros		Realización de predicciones			

¹⁶¹ Ibid. p.79.

¹⁶² FOURIE, I. Should we take disintermediation seriously? *The Electronic Library*, vol.17, n°1, January 1999, pp. 10-12.

¹⁶³ Ibid. pp.12-15.

desarrollos que aclaren vía a seguir		negativas.			
Defensa que son ellos quienes tienen las habilidades para tareas exclusivas.					

Tabla 2. Actitud del bibliotecario ante la desintermediación.

Las primeras tres reacciones son poco prometedoras para el futuro de los mediadores. A menudo las reacciones pasivas se pueden atribuir a la ignorancia, pero lamentarse por ella no prevendrá sus consecuencias. En lugar de ello, los documentalistas deben estar alertas al entorno cambiante y a las expectativas de los usuarios como prerrequisito para su supervivencia en el contexto de la Sociedad de la Información. Los cambios derivados del entorno general, como la educación o el entorno laboral de los usuarios finales, también deben ser tenidos en cuenta.

El proclamarse único experto en gestión de la información podría tener consecuencias fatales. El profesional duda de las habilidades del usuario final para buscar y evaluar con efectividad los resultados de su búsqueda, pero ningún ataque contra las búsquedas de usuario final los disuadirá de ello si verdaderamente ese es su deseo. Por el contrario si reciben un servicio excelente por parte de profesionales, como puedan ser servicios de valor añadido, pueden convencerse de la importancia del documentalista. Lo mismo puede decirse de los programas de formación para usuarios finales competentes, que demuestran el poder de una búsqueda efectiva. En otras palabras, hemos de demostrar con hechos nuestra capacidad. En lugar de discutir con usuarios finales y otros agentes implicados que podrían amenazar la posición de los mediadores, tiene más sentido dedicar el tiempo y la

energía a afianzar la posición de los documentalistas.

Muchos autores opinan que los documentalistas son profesionales polifacéticos que necesitan de un conjunto más amplio de competencias que no sólo surgen del campo de la Documentación¹⁶⁴. Entre las opciones estarían webmasters, editores, gestores del conocimiento o diseñadores de intranets¹⁶⁵. Sin embargo, sería desastroso si estos profesionales abandonan su experiencia en organización y recuperación de información, diseño de sistemas o evaluación de las necesidades de los usuarios en los servicios de información clásico. Así, por ejemplo, localizar información en el entramado desestructurado de Internet puede ser caótico; aquí el documentalista puede ayudar a organizar la información en las publicaciones web con mayor efectividad¹⁶⁶.

El área 3 trata acerca de las opiniones personales. Aunque a menudo se cite la opinión de los expertos, siempre debe tenerse en cuenta el punto de vista personal del usuario acerca de la desintermediación. El conocimiento de los expertos siempre es un valioso punto de referencia, pero debemos recordar que no siempre es el más adecuado para el usuario. Tenopir, si bien predijo que la búsqueda directa del usuario final transformaría por completo el papel del mediador, actualmente mantiene que existe un futuro para ambas situaciones: a) la búsqueda directa del usuario final y, b) la búsqueda a través de intermediarios. Si bien es esencial conocer las opiniones autorizadas, nunca debe ser el único fundamento para la toma de decisiones¹⁶⁷.

En el área 4 se propone el empleo de escenarios y previsiones futuras. A partir de este modelo es posible destacar algunos aspectos acerca del papel del mediador, entre los cuales destacamos los siguientes:

- a) Los gestores y los técnicos tienen una visión distinta del futuro del documentalista,
- b) Hay dudas sobre cómo los nuevos papeles para los mediadores podrían tener

¹⁶⁴ MOREIRO GONZÁLEZ, J.A.[et al.]. Desarrollo profesional y opinión sobre la formación recibida de los titulados universitarios en información y documentación de las universidades públicas de Madrid (2000-2005). *El Profesional de la Información*. vol.17, nº3, p.261.

¹⁶⁵ ECIA. *Euro-referencial en información y documentación. Volumen 1: competencias y aptitudes de los profesionales europeos de información y documentación*. Madrid: Sedic, 2004. En: http://www.sedic.es/p_euro-referencial.htm. Consultado el 15/12/2009.

¹⁶⁶ FOURIE, I. Op. cit. p.11.

¹⁶⁷ TENOPIR, C. Trends in end user searching. *Library Journal*. Diciembre, nº 35-36. En: <http://www.libraryjournal.com/archive/> Consultado el 15/12/2009.

sentido,

- c) Existen temores sobre el status del documentalista al dejar el control en manos de los usuarios finales,
- d) El papel del documentalista está influido por los cambios generales en el contexto cultural, social y organizacional.

En función de esto se pueden establecer cuatro escenarios para la desintermediación:

- I) la figura del mediador seguirá existiendo,
- II) el mediador podría no ser un documentalista ni un bibliotecario,
- III) el mediador tendrá un papel diferente,
- IV) el control del flujo de información puede estar restringido por una serie de factores.

Al considerar el contexto global en el que el documentalista desarrolla su trabajo, podría parecer que todos estos escenarios son viables. El escenario dependerá del tipo de servicio de información, la organización a la que pertenece, el cambio social, así como las posturas individuales.

En el área 6 se trata de otras respuestas a situaciones similares. Las empresas e instituciones también se enfrentan a un acuciante problema de desintermediación. Tras una inicial desorientación reaccionaron reclamando sus propias responsabilidades; al menos ellos sí poseían la posición financiera para apoyar sus exigencias. En muchos países, especialmente del ámbito anglosajón, la legislación introdujo al mismo tiempo reformas para consolidar sus funciones. No obstante ciertas funciones fueron abiertas a los usuarios finales, en condiciones tales que aceptasen la responsabilidad de sus propias decisiones.

Aunque se puede aprender mucho de la reacción de las empresas y organismos a la sobrecarga de información, no todas estas soluciones son aplicables a un centro de documentación. Especialmente porque no se cuenta con una financiación que permita desarrollar servicios revolucionarios. Tampoco existe un marco legislativo que proteja la figura del mediador. Una diferencia importante entre los centros de documentación y las intranets de las empresas reside en que aquellos nunca han aceptado su responsabilidad en la calidad, corrección o exhaustividad en las búsquedas. Jones observa que aunque la información incorrecta puede tener serias consecuencias, la rentabilidad y la eficacia tiende

a ser mayor en otras profesiones. Al contrario que en el caso de las plusvalías, el usuario final no es consciente del poder de la información. Es menos amenazante para el usuario final asumir la responsabilidad por las decisiones sobre sus propias necesidades de información y búsquedas que cuando está invirtiendo su dinero. Estaríamos en un escenario diferente si el usuario pagara por la información, especialmente si pensamos en las tarifas de los servicios comerciales de bases de datos¹⁶⁸.

Algunos servicios de información han compartido su experiencia y sus respuestas ante la desintermediación. Se insiste en: a) la formación de cara a asumir nuevos papeles, b) mejorar la imagen del documentalista y de los servicios que puede ofrecer, c) monitorización del comportamiento de búsqueda de los usuarios en contextos concretos, así como de la calidad de la información a la que tienen fácilmente acceso, para determinar el papel del mediador.

Finalmente el área 6 se refiere a la reflexión sobre comportamientos pasados. Dado el rápido desarrollo de las tecnologías de la información en los últimos años, existen documentalistas con una deficiente alfabetización informacional, y sobre todo desconocedores tanto de cómo gestionar información digital (bases de datos, sitios web) como de los lenguajes para la organización del conocimiento (ontologías o topic maps).

En un principio parecía que la búsqueda en línea estaba dirigida a una élite. En las bibliotecas públicas parece haberse perdido el tren para ser verdaderos productores de servicios en línea. Estudios sobre la intermediación en la búsqueda también dejan dudas sobre cómo se han adoptado, practicado y ajustado las habilidades del documentalista. Saracevic afirma que la búsqueda en línea es todavía imprecisa¹⁶⁹.

Aunque se previó una progresiva extensión de las búsquedas de usuario final y se adivinaron nuevos papeles para los mediadores parece que los documentalistas no se lo han tomado demasiado en serio. No se observan programas internos de reciclaje de los mediadores en habilidades de búsqueda, aseguramiento de la calidad y evaluación del rendimiento. Incluso han sido ampliamente ignoradas las propuestas sobre mejorar las capacidades formativas, los métodos y materiales didácticos de los documentalistas. En lugar de reconsiderar a los documentalistas como formadores en potencia, se arremetió

¹⁶⁸ JONES, J. *Intermediation and the digital library*. En: www.library.ubc.ca/jones/idl.html. Consultado el 15/12/2009.

¹⁶⁹ SARACEVIC, T. *Digital Library: toward an evolution of concepts*. En: https://docubib.uc3m.es/BIBRECOMENDADA/GESTDOC/gestdoc_06.pdf. Consultado 15/12/2009.

contra la falta de calidad de las búsquedas de los usuarios finales. La bibliografía reciente muestra sin embargo una tendencia hacia la recontextualización del documentalista como formador, sobre todo en relación con Internet¹⁷⁰.

En cuanto al pasado comportamiento de la profesión, sobre todo en su papel de mediador o intermediario, hay algunas preguntas que el documentalista debe plantearse en su carrera profesional como mediador. Preguntas relativas todas a ellas a su capacidad para manejar las nuevas tecnologías de la información y la comunicación, pero en cualquier caso no hay modo de que los documentalistas detengan un proceso a gran escala de desintermediación si no son verdaderos especialistas en su campo.

Los profesionales de la información y las organizaciones informativas forman parte de un contexto social más amplio. Su propio entorno de trabajo estará influido por la sociedad en general y por otros participantes en este campo en particular.

2.2.1 Aspectos que influyen en la desintermediación.

- I) La sociedad en su conjunto. Cada día más personas tiene acceso a ordenadores y a Internet. La preocupación por la posesión simultánea de conocimientos informáticos e informativos está creciendo entre los trabajadores así como entre los ciudadanos. Se podría pensar que el acceso a ordenadores seguirá aumentando y que más personas participarán de la excitante variedad de información existente en Internet, aunque por otro lado siguen existiendo grupos sociales con escaso acceso a la información.

Además, la sociedad está experimentando cambios apreciables en la educación: educación a distancia, enseñanza basada en recursos se están convirtiendo en realidades, puesto que son aptos para el aprendizaje a lo largo de toda la vida y la formación continuada. Estos parámetros educativos requieren que los individuos posean acceso a la información o que se les provea de ella¹⁷¹.

- II) Los procesos de búsqueda de información. Para determinar las posibilidades de

¹⁷⁰ LÓPEZ YEPES, J. El nuevo profesional de la información del conocimiento y de la comunicación: el bibliotecario universitario. *Anales de Documentación*, nº10, 2007, p.278. En: <http://revistas.um.es/documentacion/article/viewFile/1191/1241>. Consultado el 15/12/2009.

¹⁷¹ Manifiesto UNESCO /IFLA sobre la Biblioteca Escolar. En: http://www.unesco.org/webworld/libraries/manifestos/school_manifesto_es.html. Consultado el 04/12/2009.

desintermediación se debe analizar el proceso de búsqueda y recuperación de información. Muchos sistemas proclaman su condición de amigables al usuario, haciendo el proceso de búsqueda transparente al usuario final. El seguimiento de estas manifestaciones puede ayudar al documentalista a encontrar su nuevo rol mediante la mejora de los sistemas de información, los interfaces de usuario, o la formación de usuarios.

La búsqueda de información, tal y como ha sido practicada por los documentalistas, siempre ha sido algo más que la mera localización de documentos relacionados con una materia. La complejidad de los procesos y la importancia del empleo de fuentes de calidad deberían ser analizadas y compartidas con los usuarios finales, y una comprensión de estos procesos será vital para su formación.

- III) Las organizaciones. Las expectativas sobre búsqueda de información de las organizaciones y las de sus empleados pueden diferir. Algunas organizaciones podrían considerarlo parte del trabajo diario de un trabajador, incidiendo en la importancia de mantenerse competitivos. Otros podrían simplemente apoyar sistemas de búsqueda para el usuario final. En la sociedad de la información es casi imposible que una organización impida a sus trabajadores realizar algún tipo de búsqueda de información. El interés en las búsquedas de usuario final dependerá de la política organizacional, así como de otros factores y presumiblemente también de la misma naturaleza del trabajo. En muchos estudios de caso, se ha señalado que aunque algunos usuarios finales tienen acceso a información en línea, siguen prefiriendo utilizar intermediarios¹⁷².

La búsqueda de información supone tiempo y puede incluso ser improductiva. Por ello queda por ver cuántos usuarios finales preferirán tomar las riendas de sus propias búsquedas de información particulares. Los informes actuales parecen predecir que finalmente no habrá ninguno. Ya que no todas las organizaciones poseen sus propios servicios de información, un mayor esfuerzo de marketing sobre sus recursos por parte de los servicios de información existentes podría incluso ayudar al intermediario a asegurarse más clientes.

¹⁷² ORTIZ-REPISO, V. [et al.]. How researchers are using the OPAC of the Spanish Council for Scientific Research Library Network. *Electronic Library*, vol. 24, nº 2, 2006, pp. 190-211.

- IV) Los servicios de información. Hay diferentes tipos de servicios de información, lo que supone que cada uno se verá afectado de forma diferente por el escenario dibujado por la desintermediación, así que las predicciones sobre desintermediación no pueden ser asumidas sin más para todas las clases de servicios de información. Es necesario el análisis particular de cada servicio. A este nivel podría parecer que las bibliotecas públicas fueran a tener muchos problemas para asegurar su papel como mediadores, pero una acción drástica podría prevenir la desintermediación total en este campo. Parece obvio sin embargo que a menos que un servicio de información adopte activamente un programa para la reintermediación no podrá seguir manteniendo su papel actual.
- V) El usuario final. El usuario final es un sujeto individual con sus propias destrezas, preferencias y limitaciones de tiempo, dentro de las cuales debe completar sus tareas. Las necesidades y preferencias del usuario final de una biblioteca concreta deberían ser delimitadas antes de tomar cualquier decisión sobre las posibilidades de la desintermediación y su repercusión en la labor del bibliotecario.

La búsqueda en línea requiere ciertas habilidades, así como conocimiento de la disponibilidad y calidad de las fuentes de información. Para una búsqueda efectiva por parte del usuario final, la formación es esencial, y como hemos visto este es uno de las principales nuevas tareas del bibliotecario. No podemos dejar de destacar que incluso si el usuario final está en condiciones de buscar su propia información, no todos lo harán. Así como existen numerosas tareas que la gente podría hacer por su cuenta y sin embargo existen profesiones para ello; así la documentación es una profesión que efectúa búsquedas para usuarios finales que aquellos podrían realizar por su cuenta. La mayor disponibilidad de información incrementará las expectativas del usuario final en servicios de calidad y en sus demandas de formación y apoyo, pero seguirán contando con los documentalistas para la obtención de información¹⁷³.

- VI) El bibliotecario como especialista en información. Los mediadores deben considerar críticamente sus propias habilidades y capacidades para el continuo

¹⁷³ Id.

desempeño de su profesión, teniendo en cuenta la influencia de las expectativas de su profesión y especialmente la visión del futuro que sobre los mediadores posean los gestores de los servicios de información. En este sentido algunos de los nuevos roles del bibliotecario como mediador entre su centro y su comunidad de usuarios son la negociación con comercializadores de bases de datos y proveedores de servicios de información, la organización del acceso por materias a fuentes de información no estructurada, el diseño de intranets, trabajo junto con otros agentes como editores para la mejora del acceso a la información o proporcionar formación y servicios de ayuda para usuarios finales.

2.2.2 La biblioteca como intermediario experto.

En general todas las conceptualizaciones de biblioteca inciden de una u otra manera en su papel mediador entre el usuario y los recursos de información potencialmente disponibles, con la finalidad de darle un sentido de los mismos al usuario final. El término “mediador” lo empleamos en el sentido de que la biblioteca ofrece servicios al usuario sin los cuales sería incapaz de acceder y usar la información; es decir, el centro aplica servicios de valor añadido al proceso de localización y uso de recursos de información que sin ellos serían desconocidos o consumirían mucho tiempo del usuario en su localización. Esta labor de intermediación no tiene por qué ser evidente. En el caso de los topic maps supondrá un sistema integrado en el propio proceso de búsqueda adaptado a las demandas informativas de cada usuario en particular.

La terminología empleada para describir el papel intermediario resulta bastante confusa. Entre los términos alternativos que se emplean para designar esta función tenemos los siguientes¹⁷⁴:

- a) Agente. Incide en el hecho de que el centro de documentación actúa en nombre del usuario, localizando la información que pueda ser de interés y presentándola de acuerdo con un esquema preconcebido con independencia de las características de sus usuarios. Un agente es un programa que emprende acciones para un usuario particular sobre la base del conocimiento de las preferencias del usuario. Más

¹⁷⁴ BROPHY, P. Op.cit. pp. 114-116.

concretamente en el caso de los “agentes inteligentes” pueden llegar a adaptarse a los distintos entornos sobre la base de los gustos y comportamientos de sus usuarios.

- b) Guía. Enfatiza la asistencia al usuario en la elección de las estrategias de búsqueda más adaptadas a su entorno informativo, además de formar al usuario en técnicas de búsqueda.
- c) Mediador. Supone que las centros de documentación protegen a los usuarios de la inabarcable cantidad de recursos de información desestructurados por medio de la interpretación de las necesidades informativas de sus usuarios y seleccionando los materiales proporcionados.
- d) Broker. Se centra en el empleo de expertos para la localización de un documento en vez de hacerlo los usuarios. Es análogo al bróker de las inversiones, quien aconseja a los clientes e incluso es capaz de invertir por sus clientes. Aquí lo empleamos para referirnos a los programas intermediarios entre los servidores y los clientes, o bien entre los repositorios y el demandante de información. Es un concepto un poco más específico que el referido como guía.
- e) Portal. Sugiere que la biblioteca actúa como entrada a los recursos de información. El problema con este término es que los portales bibliotecarios son mucho más que meras interfaces de acceso a la información. Ahora se emplea para designar un amplio grupo de tipos diferentes de servicios.

2.2.3 La teoría de la desintermediación aplicada a la comunidad de usuarios.

Cada biblioteca sirve a un grupo de usuarios definido. Para el CEPC su comunidad de usuarios son los investigadores y alumnos de los postgrados del centro, así como técnicos de la administración. Una de las primeras tareas del centro es la definición de los criterios que ayudan a concretar el perfil tipo de los usuarios que demandarán los servicios de documentación y biblioteca del centro. El CEPC reconoce su grupo de usuarios como aquellos registrados como personal o estudiante de la institución, pero también se reconocen subgrupos de usuarios para establecer niveles de acceso a los distintos servicios del centro. El universo de posibles usuarios está constituido por el total de miembros de la

institución y de las personas que a ella acuden. La población de usuarios es el subgrupo registrado en los servicios de biblioteca. Es necesario diferenciar entre universo y población de usuarios porque el centro necesita obtener información específica acerca de las personas registradas para usar los servicios de la biblioteca, especialmente de aquellos orientados a grupos de perfil más específico¹⁷⁵. Una vez definida la población de usuarios, esta puede a su vez subdividirse de acuerdo con criterios más específicos que permitan la constante adecuación de los servicios a cada perfil de usuario.

2.2.4 La teoría de la desintermediación desde el punto de vista de la colección.

Así como existe un universo de potenciales usuarios, también existe un universo de información. Universo de información que puede entenderse como la suma de toda la información existente en nuestro centro y que es potencialmente accesible¹⁷⁶. Así si incluyen tanto los libros en papel como los libros electrónicos, las revistas en ambos formatos o las bases de datos, incluido el catálogo. Durante mucho tiempo los documentalistas han puesto toda su atención en la rápida explosión de las publicaciones impresas y las dificultades que entraña el intento por proporcionar un acceso ordenado al universo de información por parte sus usuarios. El advenimiento de la información digital, junto con el desarrollo de la información web y las redes de comunicación ha supuesto un nuevo contexto para el desarrollo de las habilidades y técnicas de los documentalistas en su labor mediadora de facilitar el acceso a la información a sus usuarios, al disponer de un interfaz de acceso que dota al usuario de gran autonomía en su interacción con el sistema. Además debemos tener en cuenta que en el universo de información digital existe documentos propietarios con un acceso restringido a un grupo de usuarios, es lo que se denomina Internet invisible y que contiene mucha más información de la que es accesible por web.

Así como las bibliotecas pueden actuar como intermediarias de información, ahora su tarea más importante, como hemos estado viendo, es organizar de alguna manera el caótico sistema informativo de la información digital. Esto lo hacen por medio de la selección y descripción de recursos de información que se ofrecen a los usuarios.

¹⁷⁵ SANZ CASADO, E. *Manual de estudios de usuarios*. Madrid : Fundación Germán Sánchez Ruipérez, 1994, p.19-20.

¹⁷⁶ BROPHY, P. Op. cit. pp. 120-121.

Durante la selección, las bibliotecas seleccionan materiales tanto en papel, de los que es propietaria la biblioteca; como en formato digital, que pueden ser propiedad o no del centro, pero que se caracterizan por tener una mínima garantía de accesibilidad y calidad informativa.

El término colección es el modo más adecuado para la descripción de fuentes seleccionadas por una biblioteca particular del universo de posibles recursos informativos; teniendo en cuenta que la biblioteca siempre ha sido un organismo con tendencia a un crecimiento orgánico de sus recursos y servicios. Esto quiere decir que los recursos digitales y las redes de comunicación lo único que han hecho ha sido incrementar la velocidad para el acceso a la información¹⁷⁷. Sin embargo, esto supone que la biblioteca debe ser capaz de incrementar sus procesos de tratamiento de la información, así como adaptarlos al carácter dinámico de la edición y difusión de los documentos digitales.

Chodorow y Lyman nos sugieren que el desarrollo de la colección se irá progresivamente transformando en lo que se llamará la continua acción directa¹⁷⁸. En los últimos años se ha puesto mucha atención, en el entorno de las bibliotecas digitales, en el mantenimiento de colecciones dinámicas, dado que la biblioteca difunde documentos cuyos contenidos se están actualizando constantemente. Esta conceptualización de la colección, como proceso vivo, por el cual se dota de un sentido a la información en sustitución del que puedan asignar sus usuarios. Esto se realiza por codificación de los contenidos y selección de aquellos que puedan ser de más interés a los distintos perfiles de usuario, labor que es una profundización en el rol tradicional del bibliotecario de selección y adquisición de documentos¹⁷⁹.

Además, el proceso de selección está relacionado con el compromiso de garantizar un nivel mínimo de calidad de los documentos. La biblioteca adquiere para su colección sólo aquellos documentos que cumplan con unos criterios de calidad, recordando que la calidad está fundamentalmente relacionada con la correcta adecuación entre un producto y los

¹⁷⁷ BROPHY, P. Op. cit. p.120.

¹⁷⁸ CHODOROW, S. ; LYMAN, P. The responsibilities of Universities. En HAWKINS, B. L. ; BATTIN, P. *The mirage of continuity: reconfiguring academic information resources for the 21st century*. Boston : Council on Library and Information Resources and Association of American Universities, 1998, pp. 63-65.

¹⁷⁹ FONFA, R. From Faculty to Librarian Materials Selection: an element in the professionalization of librarianship. En MECH, T. F. ; MCCABE, G. B. *Leadership and Academic Libraries*. Cambridge : Greenwood Press, 1998.

requisitos de los usuarios. En la biblioteca tradicional se aplican criterios tales como la reputación del editor o el empleo de documentos secundarios como resúmenes, críticas y recomendaciones de especialistas. En el entorno digital estos criterios resultan bastante menos obvios y fiables, si bien se puede acceder a recomendaciones por medio de actividades cooperativas como los foros, listas de discusión y todos los servicios propios de la Web 2.0. La dificultad está en que este tipo de recomendaciones son remotas desde el punto de vista de los usuarios, y por tanto la biblioteca está peor informada tanto acerca de sus demandas informativas como de los contextos en los que cada documento puede responder a las demandas informativas de cada usuario.

Debido a que las fuentes de información tienen una disponibilidad condicionada, en función de los derechos de autor y de acceso y uso de los mismos, la biblioteca puede escoger entre subdividir la colección accesible en función del usuario, o bien indicar condiciones bajo las cuales pueden ser consultables. La versión de la colección que el usuario individual consulta se le denomina: “colección condicional”, que cambia y evoluciona constantemente. De esta manera, la colección que se le ofrece al usuario en un momento dado es la constatación del “panorama informativo” del momento.

2.3 Teoría de la organización del conocimiento en bibliotecas digitales semánticas.

En este apartado pretendemos exponer de forma general una teoría de organización del conocimiento adaptada a las demandas que los usuarios hacen de una biblioteca digital, y más concretamente de una biblioteca digital semántica.

Creemos que las características intrínsecas, en cuanto a servicios y características de la colección, de una biblioteca digital semántica jurídica obliga a un diseño específico de un modelo de organización del conocimiento, más adaptado a la organización por materias y centrado en las necesidades de aprendizaje de una comunidad de usuarios con un lenguaje especializado y un dominio de conocimiento específico. Esta especialización, o más bien contextualización de las materias representantes de la colección digital es lo que nos va a permitir aplicar los topic maps como modelo específico para la organización contextual de objetos de información jurídica textual.

Para hablar de organización del conocimiento en bibliotecas digitales semánticas debemos previamente indicar que el concepto de análisis de materias en una colección denota el análisis de documentos en indización y clasificación¹⁸⁰. Lancaster prefiere hablar de análisis conceptual¹⁸¹, aunque lo denomine también análisis de materia.

La base de datos INSPEC¹⁸² emplea la expresión “análisis de información” para referirse tanto a la indización como a la recuperación de documentos o al uso y gestión de sistemas de información tal y como vemos en la figura siguiente referida a una búsqueda por “análisis de información” en el mapa conceptual de la base de datos. Haciendo esta misma búsqueda en el tesoro se nos indica que antes de 1973 se empleaba el término de ciencias de la información, lo que nos indica que el análisis de información es un concepto desarrollado con el tiempo para designar una amplia gama de servicios en bibliotecas y servicios de información.

¹⁸⁰ HJORLAND, B. *Information seeking and subject representation. An activity-theoretical approach to information science*. London : Greenwood Press, 1997, p.39.

¹⁸¹ LANCASTER, F.W. *Indexing and abstracting in theory and practice*. London : The library Association, 1991, p.10.

¹⁸² INSPEC. Base de datos de ingeniería, informática y física. Consultado en la biblioteca de la Universidad Carlos III. Consultado el 15/12/2009.

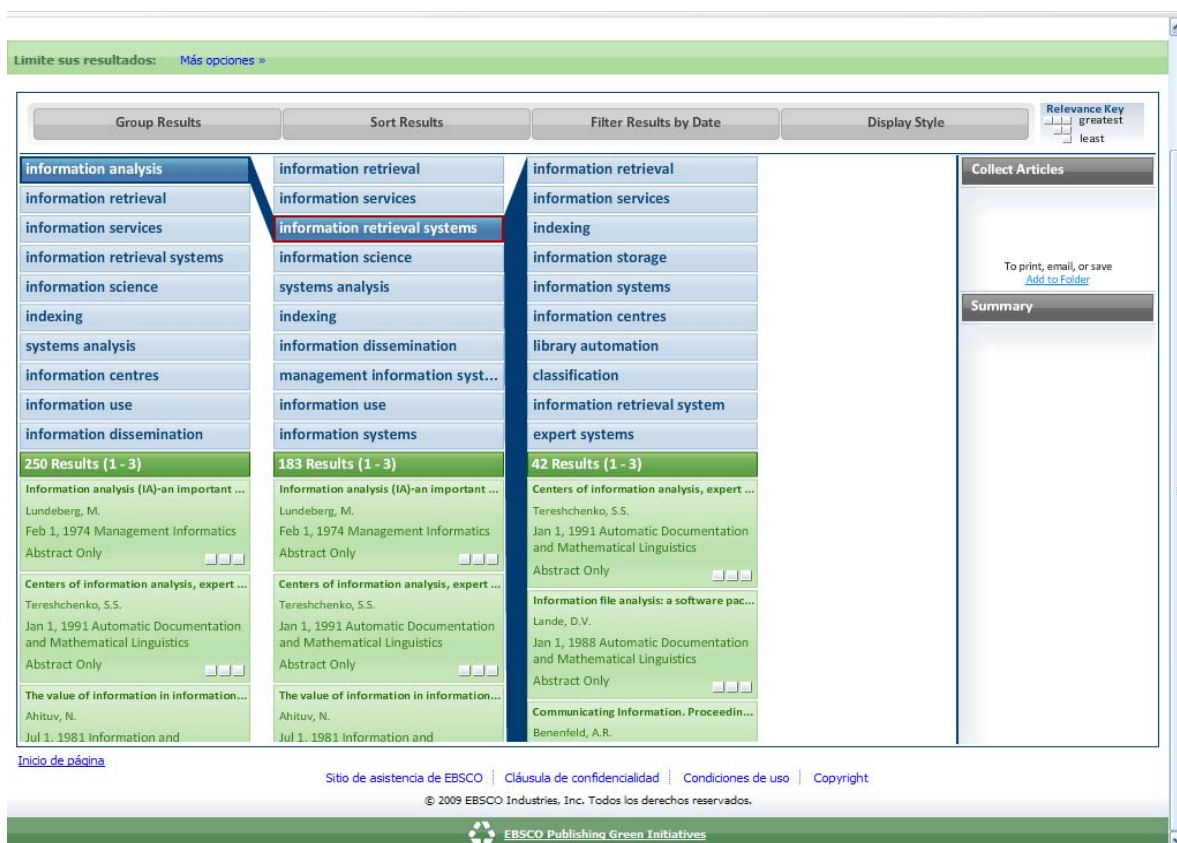


Figura. 7. Mapa conceptual de búsquedas en la base de datos INSPEC para “Information Analysis”.

La expresión “aboutness analysis” (análisis de contenido) ha sido bastante empleada en el campo de la organización del conocimiento para la planificación y organización de sistemas de información, pero particularmente en el campo de la Comunicación, la Documentación y la Lingüística computacional¹⁸³. Concretamente en el campo de la Documentación, ciertos procedimientos del tratamiento de la información documental presentan tales analogías con una parte de las técnicas de análisis de contenido que conviene aproximarlos para diferenciarlos mejor tal y como podemos constatar en los resultados de búsqueda por “análisis de contenido” de la base de datos INSPEC de la figura 3. Todo ello siempre con un mismo fin, despejar la especificidad y el campo de acción del análisis de contenido. Si se priva al análisis de contenido de su función de inferencia, si se limitan sus posibilidades técnicas sólo al análisis categorial o temático entonces efectivamente se le puede identificar con el análisis documental.

¹⁸³ BARDIN, L. *El análisis de contenido*. Madrid : Akal, 1986.

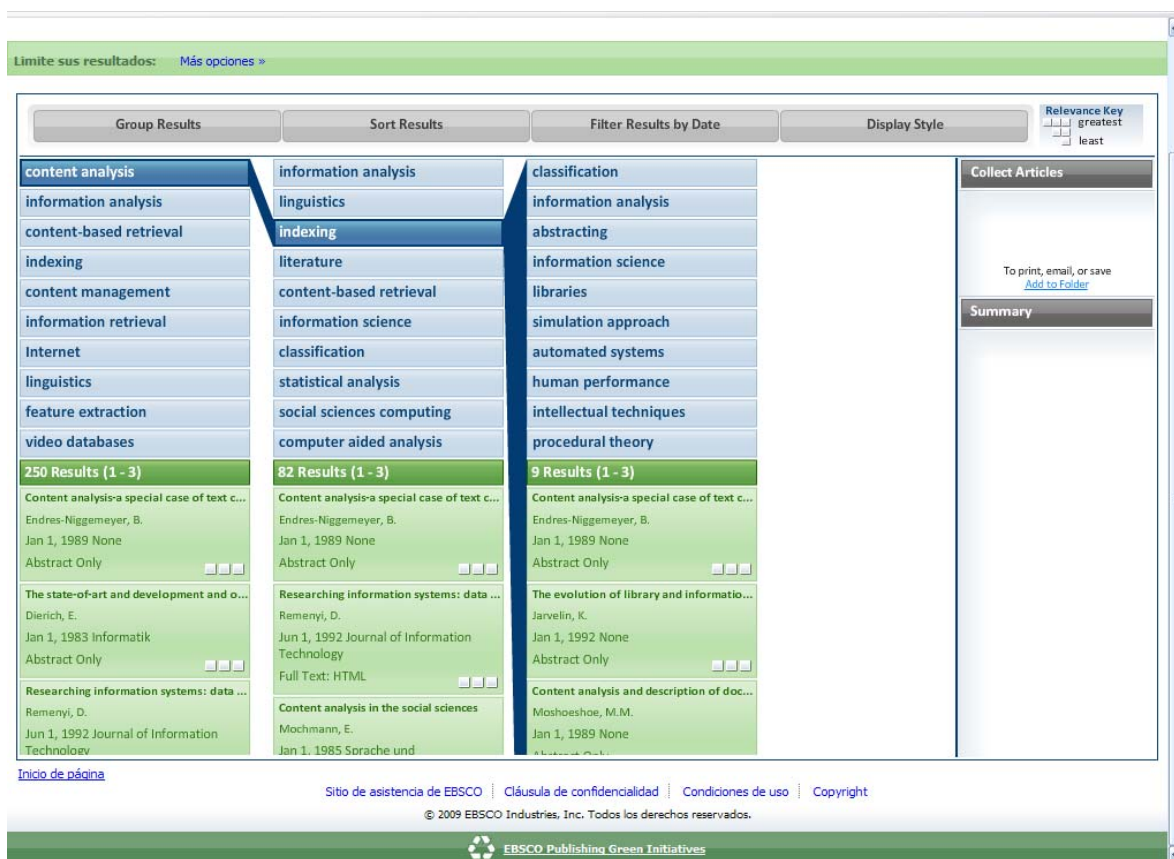


Figura. 8. Mapa conceptual de búsquedas en la base de datos INSPEC para “Content Analysis”.

Para Chaumier el análisis de contenido se le puede definir como operación o conjunto de operaciones tendente a representar el contenido de un documento bajo una forma diferente de la suya original a fin de facilitar su consulta o utilización en un estudio ulterior¹⁸⁴.

El tratamiento de la información contenida en los documentos reunidos tiende a darle una forma, a representarla de otro modo por procedimientos de transformación. Para Chaumier el objetivo de estos procesos de transformación es facilitar el acceso al usuario para que pueda obtener el máximo de información (aspecto cuantitativo) con el máximo de pertinencia (aspecto cualitativo).

Es esta concepción del análisis de contenido, conceptualizada casi como sinonímica al análisis documental la que vamos a emplear como fase previa en el diseño de nuestra biblioteca digital semántica jurídica, especialmente en tanto en cuanto aplicamos un modelo, topic maps, basado en el análisis y contextualización de las materias.

¹⁸⁴ CHAUMIER, J. *Les techniques documentaires*. París : PUF, 1974. p. 124

En esta etapa previa de análisis documental, que en nuestra biblioteca del CEPC ya se encuentra realizada por el servicio de biblioteca, se representa un documento primario con un documento secundario por medio de un metaíndice de topics. Sin embargo, podemos partir del proceso de indexación realizado en el CEPC o incluso otros organismos como el centro de documentación del Senado, efectuado con los descriptores del tesauro EUROVOC, para clasificar los elementos de información de los documentos. La ventaja de esta indización es que ya está regulada según una elección de términos adaptada al sistema y a la finalidad de la documentación en cuestión.

Los *topics*, al permitir la distribución de la información bajo una entrada que sirve de pista, son las categorías de un topic map bajo el cual se agrupan todos los documentos de una colección ubicua de documentos de acuerdo con las analogías de su contenido.

La operación intelectual de distribución en topics según el criterio de la analogía, o la representación bajo forma condensada por indización es idéntica a la fase de tratamiento de mensajes del análisis de contenido. De hecho, en los topic maps concurren procedimientos del análisis de contenido y el análisis documental que antes eran diferencias esenciales:

La documentación actúa sobre documentos; el análisis de contenido sobre mensajes. Pero en los topic maps se representan documentos con mensajes como veremos más adelante¹⁸⁵. Ej: “David nació en Ourense”, representa una fotografía de Ourense en los años 70.

El análisis documental se hace principalmente por clasificación; el análisis categorial temático es una más de las distintas técnicas de análisis de contenido que puede ser perfectamente asimilado a la asignación de clases en una ontología o de topics en un topic map¹⁸⁶.

El objetivo del análisis documental es la representación condensada de información para almacenamiento y consulta; el del análisis de contenido es el tratamiento de mensajes para actualizar indicadores que permitan inferir de una realidad otra diferente al mensaje, pero que sin embargo podemos utilizar para reformular las consultas en el sistema de recuperación en el topic map; dado que el potencial del topic map es la capacidad de

¹⁸⁵ PEPPER, S. *The TAO of topic maps*. En: http://www.ontopia.net/topic_maps/materials/tao.html. Consultado el 15/12/2009.

¹⁸⁶ BROUGHTON, V. *Essential thesaurus construction*. London : Facet Publishing, 2006. p. 135.

manejar relaciones sintagmáticas para representar relaciones entre documentos y así poder recuperarlos de forma contextualizada¹⁸⁷.

Para Hjørland la diferencia entre el análisis de texto y de contenido está en el propósito de cada uno de ellos. Para éste tiene la finalidad de la recuperación de información, en tanto para aquel implica una evaluación de la relevancia en relación con un objetivo específico. El análisis de materia es una forma de análisis de contenido que supone un proceso manual o automatizado de análisis de las materias de un documento y la consiguiente expresión de este análisis como datos representativos de la materia.

En el campo de la lingüística, se ha empleado el análisis de texto, en tanto que el análisis de contenido es más empleado en las ciencias sociales. Pero la expresión más empleada en Documentación para indicar las materias de una colección de documentos es el “análisis de materias”¹⁸⁸.

Una teoría de análisis de materia necesita de una teoría acerca de lo que es una materia, para lo que nos vamos a apoyar en las teorías de Langridge¹⁸⁹. Este nos indica que el análisis de contenido es demasiado genérico y opta por hablar del término más específico de análisis de materia. Entiende la materia de un documento como una correspondencia entre el contenido entre el documento y una serie de categorías de conocimiento preestablecidas. Categorías caracterizadas por su permanencia y su inherencia al conocimiento que designan. No obstante no tenía en cuenta la diferencia entre la indización orientada al contenido y la orientada a las demandas informativas del usuario. Además no tiene en cuenta el uso potencial del documento en el análisis de una materia, aspecto éste que ha cobrado protagonismo con las nuevas formas de documento y su difusión por internet. No obstante es interesante para introducir un marco epistemológico de topic maps la idea de categoría no permanente de conocimiento pero sí relativamente estable.

Las teorías de la indización automatizada están orientadas hacia un idealismo objetivo por el cual las categorías de conocimiento son permanentes y no sufren actualizaciones. De esta manera se está asumiendo que en la superficie de un documento se oculta una categoría de

¹⁸⁷ MOREIRO, J.A. ; SÁNCHEZ CUADRADO, S. ; MORATO, J. *Panorámica y tendencias en topic maps*. En: <http://www.hipertext.net/web/pag229.htm#Navegabilidad%20e%20inferencia>. Consultado el 15/12/2009.

¹⁸⁸ HJØRLAND, B. Op. cit. p. 39.

¹⁸⁹ LANGRIDGE, D. W. *Subject analysis: Principles and procedures*. London : Bowker-Saur, 1989, pp.5-6

conocimiento que puede ser explícita. Más que dependiente del uso potencial del documento, se relaciona con las características inherentes y permanentes del conocimiento.

Soergel plantea una teoría de la indización que relaciona la materia con su contexto¹⁹⁰, sugiere que el análisis de materias se efectúa mejor por medio de la comparación de un documento con las estructuras del conocimiento subjetivo. Es común entender la materia como la intención significativa de un autor, y por tanto reflejo de las estructuras subjetivas del mismo. Este método lo emplean los representantes de la hermenéutica clásica y es defendido por la corriente cognitiva de la Documentación por otros autores como Ingwersen¹⁹¹.

Hjørland intenta encontrar la clave del análisis de materias y el concepto de materia en la subjetividad del usuario y sus necesidades individuales. Defiende que la materia de un documento es su potencial informativo¹⁹². Potencial que no proviene de las características permanentes del conocimiento o por los intereses subjetivos del individuo o por las necesidades informativas de los autores y usuarios, sino de un análisis del dominio de conocimiento basado en la filosofía de la ciencia, la sociología y las comunidades discursivas.

Un análisis de materia implica una interpretación del potencial del documento en relación a los intereses de conocimiento de un sistema de información dado, y este análisis se efectúa en un contexto sociocultural concreto.

El análisis de contenido puede orientarse a un objetivo o bien ser general, lo cual determinará el tipo de búsqueda que pueda llevarse a cabo: por navegación o por motores búsqueda. En un servicio de información específico de un dominio, como pueda ser el jurídico, se analizan los potenciales más específicos en relación con las necesidades del servicio específico del correspondiente centro de documentación jurídico.

Soergel, nos indica, con relación a una base de datos en la que se recogen descripciones de documentos de un dominio de conocimiento concreto, que las descripciones de sus registros tienden a ser más específicas en un servicio de información general, donde las descripciones tienden a ser más descriptivas de la forma documental que del contenido de

¹⁹⁰ SOERGEL, D. *Organizing information: Principles of data base and retrieval systems*. London : Academic Press, 1985.

¹⁹¹ INGWERSEN, P. *The turn : integration of information seeking and retrieval in context*. Dordrecht : Springer, 2005. p. 35-40

¹⁹² HJORLAND, B. Op. Cit. p.40.

los documentos¹⁹³. Así podemos ver cómo el contexto pragmático juega un importante papel en el análisis práctico de las materias.

Un claro ejemplo de ello lo tenemos en la base de datos IBERLEX, donde es posible efectuar las búsquedas por tipo de disposición, lo cual es una descripción bastante sencilla y objetiva, pero es poco informativa para los usuarios que buscan documentos específicos de una materia. Un análisis jurídico debe facilitar al usuario la elección de sus necesidades y preferencias. Este objetivo se logra integrando distintos criterios de selección, así es como introduciendo los criterios de título y fecha el usuario es capaz de delimitar notablemente sus búsquedas; pero incluso con la inserción de un campo de palabras clave resulta difícil para el usuario la recuperación de documentos correspondientes a un dominio de conocimiento. El principio de análisis de materias como interpretación del potencial informativo de los documentos se modifica por factores pragmáticos. Esta modificación es a menudo tan específica que puede hacer difícil ver que una modificación del potencial de los documentos ha sido llevada a cabo. La conexión entre las necesidades de los usuarios y el análisis documental es muy indirecta, ya que es el usuario quién por sí mismo debe identificar los documentos relevantes sobre la base de criterios como el título y esto no siempre es efectivo.

El análisis de materias implica la construcción conceptual de los contenidos y de su potencial informativo. La construcción conceptual puede hacerse relacionando acontecimientos, personas, lugares, tiempos, pero el resultado puede ser tan sólo relevante para algunos tipos de necesidades informativas. Otra forma de construcción conceptual es relacionando conceptos abstractos, lo que implica la realización de generalizaciones basado en el análisis teórico y la abstracción. En la práctica no hay nada que impida la asignación de muchas materias a un mismo documento, tanto positivista con materias concretas como por medio de generalizaciones con conceptos abstractos. Las descripciones con materias abstractas o generalizadas sirven a necesidades de conocimiento más esotérico. Un análisis de materias no puede ser concebido separadamente de su entorno y el contexto en el cual se efectúa, incluyendo el marco cognitivo individual o colectivo sobre el que se fundamenta el análisis. Pero el objetivo del análisis no debe ser subjetivo o influido por la persona que lleva a cabo el análisis. Más bien, debe ser tan preciso y objetivo como sea posible, así como amoldarse tanto como sea posible a su objeto de análisis.

¹⁹³ Ibid. p. 45ss.

Hjørland nos dice que no es posible un análisis de materias correcto para un documento, en el sentido que no hay “respuestas correctas”¹⁹⁴. El análisis documental no es cuestión de la localización de una correspondencia entre un documento y las posibles asignaciones de materias. El criterio de calidad decisivo en el análisis de las materias de los documentos está en que produce datos que representan materias, lo cual refleja más las características esenciales que incidentales del documento. Tal evaluación de las características esenciales asume que existe un análisis independiente en relación con la autodescripción del propio documento a través del título o los encabezamientos. No existen reglas para el análisis de las características esenciales de los documentos. En ocasiones la metodología de análisis de un documento constituye una característica esencial, pero en otras ocasiones no es tan importante.

Podemos afirmar que el análisis de materias asume la imparcialidad, una capacidad para ver a través de posibles aspectos de autopromoción comercial de un documento. Así, por ejemplo, es común que un documento demande tratar un topic de manera bastante más general de lo que realmente hace.

Otra cuestión es que el análisis de materia debe emprenderse desde el punto de vista de los intereses de conocimiento profesional, desde el punto de vista de distintas disciplinas y subdisciplinas. En un sistema universal la idea es situar un documento en un único contexto temático. Esto supone problemas irresolubles en el momento de determinar si un documento de filosofía política es acerca de filosofía o de política. Si empleamos un tesauro o una clasificación este problema desaparece: cada disciplina decide sobre la relevancia de cada documento y le asigna un lugar en el contexto de esa misma disciplina. Un documento tiene no sólo una materia verdadera sino que tiene muchas potencialidades epistemológicas a las que se les da prioridad sobre la base de los puntos de vista de la disciplina. Los registros bibliográficos pueden contener distintas prioridades disciplinares, pero debe ser posible diferenciarlas. Los descriptores empleados en la base de datos IBERLEX, tomados de los tesauros de Derecho, se localizan en el campo palabra-clave de cada uno de los registros de la base de datos y por ello no son concebidos como ruido por una persona que esté buscando obras de filosofía.

En la práctica, el análisis de contenido lo realizan los documentalistas a partir de un lenguaje controlado con el objeto de poder representar el resultado del análisis de forma

¹⁹⁴ Ibid. p.41.

coherente. Tanto los lenguajes precoordinados como los postcoordinados son lenguajes para la indización y recuperación de información¹⁹⁵. Los documentos analizados con cualquiera de este tipo de lenguaje, representan el conocimiento organizado. Sin embargo estos lenguajes pueden afectar al modo en que se efectúa la asignación de las materias por la propia concepción teórica del lenguaje. Un lenguaje controlado es en sí mismo un sistema de asistencia al análisis de contenido, pero no siempre es el más adecuado para cada tipo de documento.

Su ventaja reside en que el análisis de contenido que se realiza regularmente con estos lenguajes nos garantiza unos niveles máximos de consumo de tiempo por parte de los usuarios durante la realización de las búsquedas al mismo tiempo que se establece una coherencia en los resultados de búsqueda¹⁹⁶.

Entre las desventajas se encuentran que la estructura de los lenguajes controlados puede llegar a suplantar la propia estructura de los documentos indizados sin que haya una adaptación del lenguaje al esquema del documento. Así si un sistema de recuperación de información otorga prioridad a unas facetas frente a otras que no fueron incluidas en su momento pero que pueden ser las más adecuadas para el análisis de contenido del documento, lo cual lleva en ocasiones a la realización de un análisis de contenido inadecuado. Es más, en muchas ocasiones el documentalista ni siquiera es consciente de las limitaciones del sistema que permitirían su adecuación a las facetas o puntos de vista expresados en los documentos.

Es importante hacer una distinción clara entre el análisis de contenido, la interpretación del potencial del documento, y por otro lado el posterior proceso de traslación del contenido del documento a las materias de un lenguaje controlado. Si se solapan ambos procesos es imposible la evaluación del análisis de contenido efectuado sobre el documento y del lenguaje controlado de forma independiente, lo que puede llevar a entender la clasificación e indización como métodos de análisis diferentes. Langridge nos dice que en la práctica el análisis de contenido crea bastantes más problemas que el proceso de traslación de los descriptores empleados en el análisis de contenido a las materias de un sistema de clasificación¹⁹⁷, razón por la cual optamos por un análisis de materias multirepresentacional.

¹⁹⁵ GIL URDICAÍN, B. *Manual de lenguajes documentales*. Gijón : Trea, 2004, p. 13.

¹⁹⁶ *Ibid.* p.16

¹⁹⁷ *Ibid.* p.12

El concepto de polirepresentación fue planteado por uno de los principales representantes de la corriente cognitiva en ciencias de la información, Ingwersen, a mediados de los años 90. El propósito de su planteamiento era que durante los procesos de búsqueda de información existen categorías o topics pertinentes a la demanda informativa del usuario hasta un cierto límite a partir del cual dejan de serlo¹⁹⁸. Y en efecto, en ocasiones es importante limitar el significado de un término a un contexto particular; y en otras ocasiones resulta útil el empleo de un topic en disciplinas con unos límites conceptuales difusos como puedan ser entre la Documentación y la Biblioteconomía. A pesar de los posibles problemas de homonimia que puedan surgir con las búsquedas en texto libre es un hecho que sobre todo para las búsquedas de objetos de información multimedia resultan de más valor. Una clara aplicación de este principio sería la ejecución de un análisis documental con verbos dada su capacidad para interrelacionar topics de distintos dominios¹⁹⁹.

En verdad el concepto de polirepresentación no es más que reconocer en la recuperación de documentos el principio de incertidumbre en la ciencia y más especialmente cuando se trata de ciencias sociales y humanidades. La orientación semánticamente restrictiva de los lenguajes documentales es muy adecuada cuando se trata de aplicarlos a unas colecciones documentales bien delimitadas pero si estamos hablando de una colección ubicua, constituida por objetos de información en muy distintos formatos y hospedados en servidores de cualquier lugar del mundo, servidores que en muchas ocasiones contienen documentos con propósitos muy diferentes los unos de los otros, entonces la capacidad de asistencia a la recuperación de los lenguajes documentales conocidos se vuelve muy limitada a causa del elevado grado de incertidumbre e impredecibilidad que existe acerca de las demandas informativas de una comunidad de usuarios universal²⁰⁰. Y una de las incertidumbres que más afectan es la relativa a conocer las materias o topics más adecuadas para la resolución de un problema y hasta qué punto su validez en un campo de conocimiento lo es también en uno afín. Esto implica que los actuales diseños de sistemas de información basados en uno o dos lenguajes documentales resultan muy limitados por representar una visión lineal y única de la realidad que se representa.

¹⁹⁸ INGWERSEN, P. Polyrepresentation of information needs and semantic entities: Elements of a cognitive theory for information retrieval interaction. En *ACM/SIGIR-94*. London : Springer, 1994, pp. 101-110.

¹⁹⁹ MOREIRO GONZÁLEZ, J.A.[et al.]. Metodología del proyecto REID : Desarrollo de un sistema de recuperación para entornos de información dinámica: Tesoros de verbos, implementación del estándar ISO/ICE 13250:1999. En *Jornadas de Tratamiento y Recuperación de Información*. En: <http://www.fundacion.uc3m.es/jotri2003/ponencias/metodologia.pdf> . Consultado el 15/12/2009.

²⁰⁰ INGWERSEN, P. ; JÄRVELIN, K. *The Turn. Integration of Information Seeking and Retrieval in Context*. Dordrecht : Springer, 2005, p. 206.

Para una biblioteca digital con una comunidad de usuarios virtual y heterogénea se hace imprescindible contar con un modelo de información que ofrezca distintos puntos de acceso que se complementen entre sí con una coordinación lógica. A esto lo denominaremos la polirepresentación del análisis de materias.

Por otra parte el concepto de búsqueda interdisciplinar hace más evidente la necesidad de la polirepresentación para no limitar el análisis a un único dominio de conocimiento, y además no tener que sustituir un análisis de materia demasiado específico para ser pertinente por otro con una clasificación alternativa reutilizada para nuestro dominio²⁰¹.

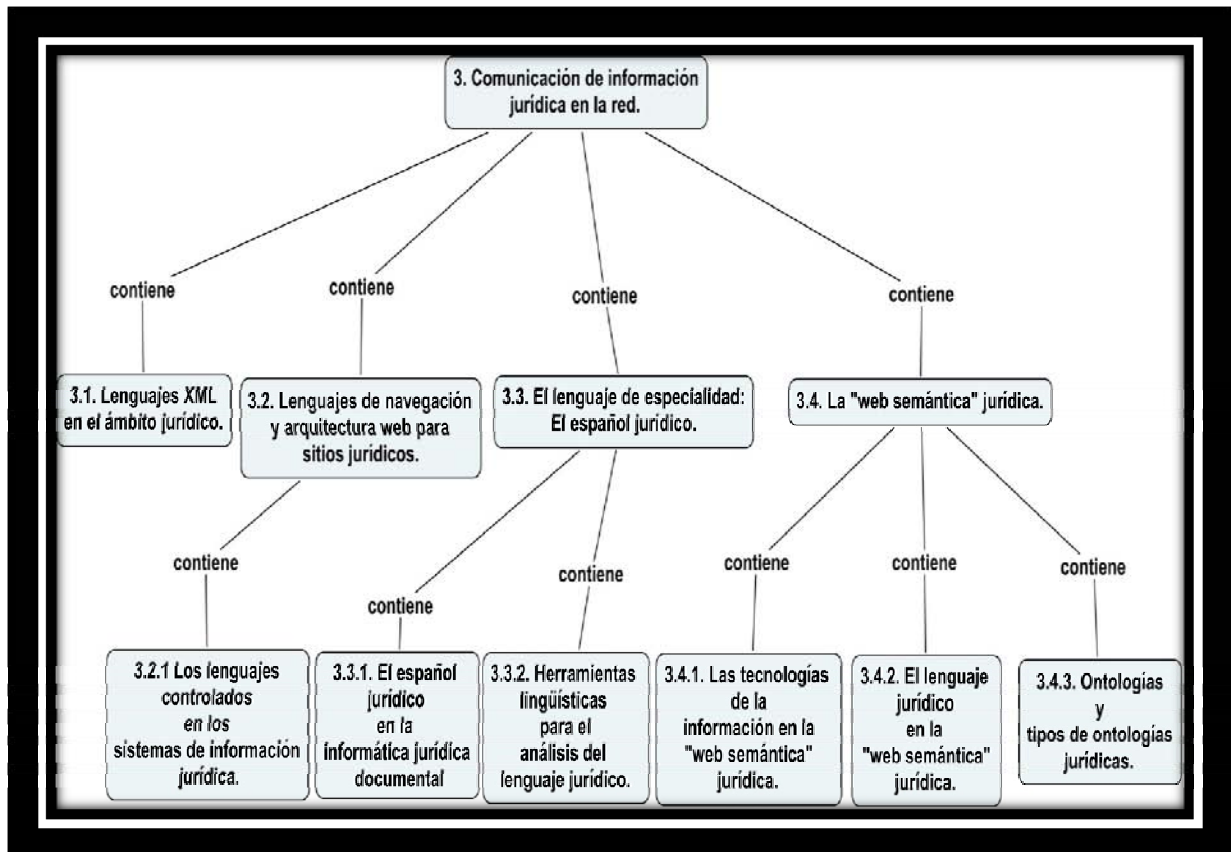
Los distintos dominios de conocimiento podrían ser entendidos como distintos modos de organizar una misma realidad en lo que sería un tipo de polirepresentación. El derecho, la filosofía, la ciencia política pueden estudiar un mismo fenómeno como por ejemplo la libertad, pero desde distintos puntos de vista e integrando el concepto en distintas estructuras conceptuales.

Finalmente debemos tener en cuenta que en el conocimiento la interdisciplinaridad puede verse condicionada por las diferentes percepciones de los seres humanos, la sociedad, la ideología o sencillamente la experiencia personal, lo cual hace más evidente la necesidad de que nuestros modelos de información sean capaces de proporcionar polirepresentación al análisis de materias. Y como veremos a lo largo de esta tesis pensamos que es el modelo *Topic map* el más adecuado para aplicar el concepto de polirepresentación a la organización y recuperación de objetos de información en una biblioteca digital semántica.

²⁰¹ KLEIN, J. T. *Interdisciplinarity: History, theory and practice*. Detroit : Wayne State University Press, 1990.

CAPÍTULO 3.

Comunicación de información jurídica en la red.



Introducción.

En este capítulo vamos a exponer la interrelación de los distintos lenguajes que participan en la comunicación de información jurídica por internet. Nos centramos en los lenguajes, por considerarlos una parte esencial del diseño de una arquitectura de información con topic maps, adecuada a las características de una biblioteca digital jurídica.

Tal y como vemos en la figura siguiente, en la comunicación de información jurídica ejercen un papel central la combinación de tres tipos de lenguajes, si bien cada uno de ellos por separado ejerce su papel a cada nivel de la arquitectura de información. Estos lenguajes son:

Lenguajes Web, concretamente los lenguajes XML creados para el ámbito jurídico. Destacamos LegalXML, creado para facilitar la interoperabilidad entre sistemas de información jurídica como veremos más adelante.

Lenguajes para navegación en sitios web como portales o bibliotecas digitales. Son vocabularios controlados no sólo para asistir al usuario en la recuperación de información sino que permiten al usuario interactuar con los recursos informativos a los que proporciona acceso la biblioteca del CEPC.

Lenguajes de Especialidad, y para nuestro caso concreto el lenguaje jurídico por ser un lenguaje con unas propiedades específicas propias del dominio jurídico que deben ser tenidas en cuenta para el funcionamiento exhaustivo y preciso del sistema de recuperación de información para las demandas informativas de nuestra comunidad de usuarios en la biblioteca del CEPC.

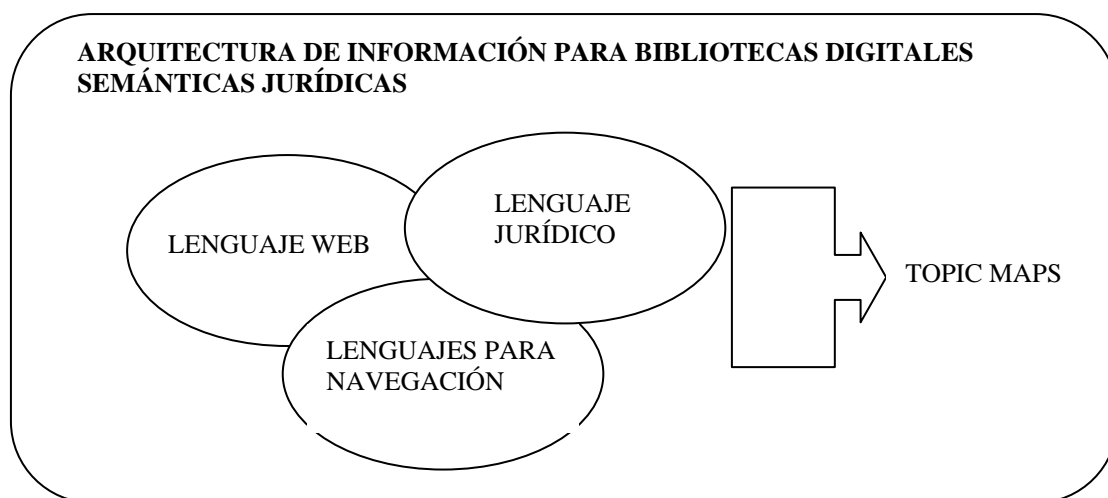


Figura 9. Esquema representativo de la relación de los distintos lenguajes en el marco de la arquitectura de información.

3.1. Lenguajes XML en el ámbito jurídico.

La estructura intrínseca de la documentación jurídica ha mejorado la efectividad de los sistemas de recuperación de información tradicionales mediante la inserción de diversos campos de búsqueda propios de este tipo de documentación. Un caso paradigmático de

esta aproximación fue la base de datos CELEX de la Unión Europea, diseñada con diez índices principales y ochenta campos²⁰².

Las marcas se han utilizado de un modo simple como ayuda al procesado informático ya que permiten la separación de campos de una base de datos; pero con el tiempo dieron paso a sistemas más complejos como los procesadores de texto que permiten formatear un texto para su visualización o impresión y finalmente han evolucionado a lenguajes de marcas que posibilitan el uso del marcado de los documentos con fines documentales.

Un lenguaje de marcado viene condicionado por un conjunto de reglas que permiten fijar el tipo de marcas que se utilizarán, las marcas permitidas en cada una de las partes del documento, la forma de distinguir el texto del documento de las marcas, así como la gramática que rige el empleo de las mismas.

Es posible utilizar el marcado de documentos con propósitos documentales permitiendo separar elementos estructurales del documento o realizar abstracciones para extraer metainformación del mismo. Se distinguen así dos tipos de lenguajes de marcado: los orientados a la presentación de documentos y aquellos dirigidos a la descripción formal y de contenido de los mismos²⁰³.

A finales de los años sesenta Goldfarb recibió el encargo de diseñar un sistema de edición, almacenamiento, búsqueda y gestión de documentos legales construyendo un sistema de formateo estructural al que se denominó GML²⁰⁴. En 1986 se convertía en un estándar, SGML, metalenguaje de etiquetado de texto convertido en norma ISO 8879.

Las particularidades de SGML frente a otros lenguajes de marcas²⁰⁵ pueden cifrarse en que se trata de un lenguaje no propietario, independiente de aplicaciones y plataformas, posibilita la aplicación de un riguroso control sobre el contenido de los documentos, permite la independencia entre contenido, estructura y presentación y asienta el concepto de DTD empleado para definir un tipo de documento estableciendo su estructura, los

²⁰² SCHWEIGHOFER, E. The revolution in legal information retrieval or the Empire strikes back. *Journal of Information, Law and Technology (JILT)*. En: <http://elj.warwick.ac.uk/jilt/99-1/schweigh.html>. Consultado el 15/12/2009.

²⁰³ MARTÍN GALÁN, B. ; RODRÍGUEZ MATEOS, D. Estructuración de la información mediante XML: un nuevo reto para la gestión documental. En *VII Jornadas Españolas de Documentación*. Bilbao: Universidad del País Vasco, 2000, p.118.

²⁰⁴ GOLDFARB, CH. *The SGML handbook*. Oxford : Clarendon Press, 1995, pp.14ss.

²⁰⁵ Ibid. p.23.

elementos de los que va a constar, el orden de los mismos, los valores que deben recoger, la posibilidad de ser repetidos y los elementos que pueden contener a otros.

A pesar del potencial que ofrece SGML en cuanto a recuperación de información, su complejidad hace muy costosa la utilización en sistemas de información. Por ello, ha quedado relegado a la publicación, gestión e intercambio de documentos electrónicos en grandes instituciones.

Por esta razón, se ideó un lenguaje de marcado más sencillo denominado HTML, que no deja de ser una aplicación del lenguaje SGML desarrollado por Berners-Lee. Con este lenguaje se indica cómo se deben codificar los documentos para su distribución en la web. Es el lenguaje con mayor presencia en la red y todo ello a pesar de sus notorias limitaciones: carencia de sintaxis, escasas posibilidades de estructurar la información, o insuficiente potencialidad para la recuperación de la misma.

Se han diseñado alternativas para resolver estos problemas con nuevos lenguajes como DHTML o el metalenguaje XML; en cualquier caso, desde el punto de vista de la recuperación de información jurídica, el desarrollo del hipertexto se erigió en una tecnología imprescindible para proporcionar una solución adecuada al problema de las relaciones entre los documentos jurídicos²⁰⁶.

En 1996 comenzó su andadura XML con el respaldo del consorcio W3C a fin de diseñar un lenguaje de marcas atendiendo a los distintos tipos de documentos²⁰⁷; el hecho de seguir la tendencia actual en el mundo de la programación y estar orientado a objetos incrementa la eficiencia en la interconexión de sistemas dado que facilita la interdependencia de los datos respecto de las aplicaciones. Las etiquetas XML marcan la estructura y la semántica del documento, aportando un carácter modular al documento que permite organizar internamente la información como se desee. Distingue la presentación del documento de la información sobre la semántica y la estructura del mismo. Permite indicar cómo se deben formatear, imprimir y visualizar los distintos elementos de un documento a través de especificaciones XSL. Permite una gestión eficaz de los hiperenlaces empleando el estándar

²⁰⁶ ALVITE DíEZ, M^a. *Evaluación de sistemas de recuperación de información en el entorno jurídico español*. León : Universidad de León, 2004, p.214.

²⁰⁷ GOLDFARB, Ch. *Charles F. Goldfarb's XML handbook*. New Jersey : Prentice Hall, 2002, p.12

Xpath y, dentro del ámbito de la recuperación de la información, cuenta con especificaciones para facilitar la consulta en grandes colecciones a texto completo.

Si consideramos todas estas peculiaridades del lenguaje XML en los documentos jurídicos, entenderemos las ventajas que aporta la utilización de lenguajes de marcas como medio de difusión de este tipo de información en detrimento de las tradicionales bases de datos. Entre los argumentos que se citan para su aplicación podemos destacar la posibilidad de aplicarse a documentos de cualquier tamaño, la idoneidad para gestionar referencias internas y externas al propio documento en forma de hipervínculos, o la posibilidad de que los motores de búsqueda puedan recuperar los documentos obedeciendo al contenido del lenguaje de marcas²⁰⁸.

Los lenguajes XML en el ámbito del derecho se encuentran agrupados en su desarrollo bajo la iniciativa del consorcio OASIS²⁰⁹. No podemos hablar de un único lenguaje XML para el entorno jurídico sino que existe una familia de vocabularios agrupados bajo el epígrafe de LegalXML que abarcan los distintos campos del derecho. Esta familia de especificaciones son normalizados por una sección especial del consorcio OASIS con la finalidad de: “aglutinar expertos legales y técnicos en un foro común para crear normas que rijan el intercambio electrónico de datos legales²¹⁰”.

En 1998 se constituyó el Legal XML Inc. como una organización sin ánimo de lucro en la que participan tanto entidades públicas como privadas, con la intención de desarrollar estándares técnicos abiertos, no propietarios, para su aplicación en la documentación jurídica y aplicaciones relacionadas²¹¹. Se divide en diversos grupos de trabajo, atendiendo a los tipos específicos de documentos jurídicos.

En el propio sitio del Consorcio OASIS, bajo la categoría de Law&Government se encuentran estos grupos de trabajo con sus correspondientes vocabularios y cuyas características básicas sintetizamos en este cuadro resumen:

²⁰⁸ ALVITE DÍEZ, M^a. Op. cit. p. 215.

²⁰⁹ Consorcio OASIS. En: http://www.oasis-open.org/committees/tc_cat.php?cat=lawgov. Consultado el 16/12/2009.

²¹⁰ LegalXML. En: <http://www.legalxml.org/>. Consultado el 16/12/2009.

²¹¹ Id.

INICIATIVA	DESCRIPCIÓN
LegalXML Court Filing	Fue desarrollado por la comisión denominada OASIS Electronic Court Filing TC. Esta comisión desarrolla especificaciones para el empleo de XML en la creación de documentos legales y su transmisión entre los abogados y los tribunales de justicia, así como entre distintos tribunales de justicia.
LegalXML eContracts	Con este vocabulario en sintaxis XML se proporciona un lenguaje de marcado para documentos contractuales que permita su creación eficiente, mantenimiento, gestión, intercambio y publicación. La comisión responsable de la normalización de este lenguaje es el Comité Técnico de eContracts, perteneciente al consorcio OASIS.
LegalXML eNotary	<p>El ámbito de la notaría de documentos electrónicos discurre paralelamente a la notaría en documentos papel. De hecho muchos de los procesos que hoy día se efectúan en el acto de dar fe de un documento en papel son los mismos que los empleados en la validación de documentos con firmas electrónicas. Sin embargo, dada la naturaleza intangible de los documentos electrónicos existen ciertas diferencias que deben ser normalizadas para su tratamiento digital. El potencial de la sintaxis XML para fijar dotar de estructura a los documentos electrónicos ha hecho que se creara una comisión específica de normalización del vocabulario XML para la notaría con documentos digitales.</p> <p>Con esta finalidad se creó en el Consorcio OASIS una comisión, eNotary, para la definición de normas de aplicación del vocabulario XML en la actividad de las notarías respecto a documentos electrónicos.</p>
LegalXML IntJustice	<p>Uno de los desafíos más importantes en el intercambio de información entre los gobiernos y los distintos tribunales internacionales de justicia está en lograr la interoperabilidad entre sus sistemas para poder compartir datos que ayuden a la resolución de litigios internacionales.</p> <p>El Consorcio OASIS creó para ello la comisión técnica integrada de justicia (IJTC) para la aprobación de normas que posibiliten el</p>

	intercambio de información entre sistemas.
LegalXML Lawful	El Comité técnico LI-XML del Consorcio OASIS es el responsable de la normalización de un vocabulario XML de creación de un marco que permita generar documentos estructurados para procesos legales de detención. Con él se facilitan mecanismos de autenticación incluyendo el desarrollo y armonización de normas XML semejantes para el marco de este tipo de procesos.
LegalXML Legislative	Otro de los vocabularios creados por el Consorcio OASIS en Estados Unidos para su normalización en el ámbito jurídico es el aprobado por la Comisión Técnica para Documentación, Citas y Mensajes Legislativos. Este comité no sólo ha armonizado normas para el marcado en XML de documentos legislativos, sino que proporciona un sistema para proporcionar citas de documentos no legislativos como artículos de periódico estructurados con otros vocabularios XML específicos tales como NewsML. El objetivo primero es permitir al público participar en el proceso democrático mediante la puesta a disposición de la ciudadanía de un vocabulario que facilite el etiquetado, referenciado y procesado de documentos electrónicos. Es objetivo específico poder incluir documentos con referencias cruzadas por parte de las entidades gubernamentales y servicios de valor añadido para gabinetes jurídicos.
LegalXML ODR	<p>Finalmente el vocabulario XML, denominado ODR; es un lenguaje de marcado para documentación empleada en la resolución de conflictos jurídicos en línea. Se trata de facilitar al ciudadano la posibilidad de obtener un acceso marco a los tribunales mercantiles por medio de sistemas interoperables para la resolución de conflictos mercantiles sin necesidad de contar con la presencia física de las partes. Esta iniciativa ha permitido la aprobación de un conjunto de normas que facilitan hoy en día el comercio electrónico al simplificar los procesos de resolución de conflictos con el flujo de información jurídica legal por la red.</p> <p>Existe un comité técnico encargado de desarrollar y revisar los</p>

requisitos, especificaciones y schemas para ODRXML, y que asimismo lleva a cabo una labor de armonización de las DTD y schemas entre desarrolladores, proveedores, vendedores, agencias y cualquier tipo de organización implicada en la resolución de conflictos.

Tabla 3. Grupos de trabajo del Consorcio Oasis para la aplicación del lenguaje XML al ámbito jurídico.

En España la aplicación de XML al ámbito legal está más retrasada que en Estados Unidos²¹². Sin embargo, son varias las recopilaciones de documentos legislativos y jurisprudenciales que hacen uso del lenguaje de marcado con su correspondiente tratamiento hipertextual: *Norm@civil*²¹³, proyecto del área de Derecho civil de la Universidad de Gerona que incluye legislación y jurisprudencia, *Noticias Jurídicas*²¹⁴ de la editorial Bosch que contiene normativa concordada, artículos doctrinales; o el portal IUSTEL²¹⁵.

Una de las iniciativas con más difusión sobre la materia en España es LEXML²¹⁶. Se trata de una red europea que busca el intercambio automático de información legal. Esta iniciativa está orientada a poner en marcha un proceso de estandarización en el ámbito jurídico, coordinado y dirigido a lograr diferentes DTD y esquemas que permitan su “intercomunicación”. Así, en cada jurisdicción se podrá crear una comunidad con un sitio web, con una lista de distribución como punto de encuentro.

Las diferentes comunidades LEXML constituyen una red internacional que permanece en contacto a través de listas de distribución, herramientas de la web social como los Wikis y encuentros anuales. Así la red LEXML, que está promoviendo un foro de discusión integrado en el que participan los diferentes Estados integrantes de la misma y en cada uno de los idiomas en los que están trabajando en cada uno de los proyectos. También es la base de un proyecto de fuentes abiertas que desarrolla el Diccionario RDF. Este Diccionario permite la comunicación y el entendimiento entre estructuras de datos de

²¹² VAÑÓ VAÑÓ, Mª J. XML, una herramienta al servicio del buen gobierno corporativo. *Revista Aranzadi de Derecho y nuevas tecnologías*. Nº8, 2005, p. 101.

²¹³ *Norm@civil*. En: <http://civil.udg.es/normacivil/> Consultado el 15/12/2009.

²¹⁴ *NOTICIAS JURÍDICAS*. En: <http://noticias.juridicas.com/> Consultado el 15/12/2009.

²¹⁵ *IUSTEL*. En: <http://www.iustel.com/> Consultado el 15/12/2009.

²¹⁶ *LEXML*. En: <http://www.uv.es/lexml/> Consultado el 15/12/2009.

jurisdicciones diferentes. Además, se aspira al desarrollo de una fuente abierta "Office-Programms", que pueda construirse a partir de lenguaje XML y que permita que todos los documentos legales estén, desde el principio, en "lenguaje XML".

3.2 Lenguajes de navegación y arquitectura web para sitios jurídicos.

La arquitectura de información, como disciplina dedicada a la planificación y organización de la información, nació como metodología en 1962, vinculada en ese momento con los complejos dispositivos mecánicos y eléctricos de la industria del momento²¹⁷. No fue hasta 1976 cuando Richard Wurman empleó la denominación de “arquitectura de la información” para esta nueva disciplina. En distintas publicaciones Wurman delimitó el perfil y las características de una nueva problemática, metodología y perfil profesional. Un arquitecto de la información es una persona que crea el mapa o la estructura de información que permite a otros encontrar su camino personal al conocimiento²¹⁸.

La evolución práctica del término y sobre todo del ámbito de actividad profesional hacia el que se dirige es acelerada, especialmente desde la difusión de Internet y sus tecnologías asociadas con los protocolos de la web. Con tecnologías que permiten la manipulación y el almacenamiento ingente de contenido, cuando el acceso sistemático y rápido al contenido se convirtió en una necesidad en un entorno institucional en permanente descentralización. En todo caso, la visión original de la arquitectura de información continúa centrándose en la formulación de métodos que faciliten el acceso a la información²¹⁹.

Sin embargo, la irrupción del mundo de Internet mostró que la visibilidad de un contenido no venía proporcionada únicamente por la solución visual a la que este se somete, sino por cambios más profundos. Rosenfeld y Morville se plantean en 1998 posibles soluciones a un problema que en ese momento se estaba manifestando como un cambio tecnológico:

La arquitectura especifica cómo los usuarios van a encontrar la información al definir su organización, navegación, etiquetado y sistemas de búsqueda²²⁰.

²¹⁷ MORVILLE, P. A brief history of information architecture. En GILCHRIST, A. ; MAHON, B. *Information Architecture: designing information environments for purpose*. London: Facet, 2004, p. 6

²¹⁸ WURMAN, R. *Information Architects*. New York : Graphics Press, 1997, p.62.

²¹⁹ BATLEY, S. *Information Architecture for Information Professionals*. Oxford : Chandos Publishing, 2007, p.3

²²⁰ ROSENFELD, L. ; MORVILLE, P. *Information Architecture for the World Wide Web*. USA : O'Reilly, 2002, p.4

El cambio conceptual era importante, ya que en el contexto web el problema ya no solo se situaba en la representación visual del contenido, sino también en cómo se concebía su introducción en un sistema informático complejo. Organización, etiquetado y sistemas de búsqueda son conceptos que triunfan en la concepción que emerge de esos primeros momentos de Internet: la red como un enorme y sencillo espacio de recuperación de información, como un lugar no vinculado a una tecnología física específica y donde la información puede ser recuperada de forma contextual y concreta.

Desde el primer momento, la pragmática de la arquitectura tuvo que ver con tareas muy concretas en la creación web, como el diseño de interfaces, el mapeado de un sitio o sus previsiones de cambio. En los últimos años el horizonte tecnológico se ha ido haciendo más complejo con la generalización de las tecnologías de bases de datos y nuevas aplicaciones que permiten la conectividad entre sistemas. En este entorno es donde han surgido propuestas de especialización de disciplinas. En la ebullición de terminología conceptual creada desde las consultoras privadas o desde los propios ámbitos de investigación académica surgió el concepto de la “experiencia del usuario”, el diseño centrado en la percepción y uso que el usuario hace de un sitio web y sus funcionalidades:

La experiencia de usuario puede definirse como el conjunto de ideas, sensaciones y valoraciones del usuario resultado de la interacción con un producto; es resultado de los objetivos del usuario, las variables culturales y el diseño de interfaz²²¹.

Rápidamente se comprobó que se trataba de dos aspectos de una misma moneda, pero con un ángulo de enfoque muy diferente: si la experiencia del usuario se centra en la comprensión de su comportamiento, la arquitectura de la información presta atención a la comprensión del contenido²²². En esta dialéctica, arquitectura de información y experiencia del usuario son técnicas complementarias para la representación del contenido y la asignación de condiciones interactivas a éste. De esta forma si tenemos que realizar un sitio web para una biblioteca, deberemos emplear las técnicas de la arquitectura para organizar internamente los diferentes elementos que constituyen la colección, conectarlos entre sí y concebir un sistema de relaciones que permita a la máquina identificar los gustos del

²²¹ KNAPP, A. *La experiencia del usuario*. Madrid : Anaya Multimedia, 2002, pp.13-14.

²²² HURST, M. *About information architecture*. En: www.goodexperience.com/columns/040300infoarch.html, 2000. Consultado el 10/12/2009.

usuario para proponerle ítems de la colección adecuados a las necesidades informativas del usuario, con un entorno visual accesible en el que los diferentes elementos puedan ser percibidos por el mayor número posible de usuarios, de la forma más sencilla posible y de acuerdo con nuestros fines organizativos; lo que nos lleva a estar hablando de una biblioteca digital semántica, tal y como vimos en el capítulo precedente.

3.2.1. Los lenguajes controlados en los sistemas de información jurídica.

Las peculiaridades de la documentación jurídica condicionan sus propiedades en los sistemas de información jurídica: gestión de gran volumen de información, documentos digitales a texto completo²²³, exigencia de exhaustividad en los corpora legales, demanda de precisión en los documentos como indicador de su pertinencia jurídica, resolución de las complejas relaciones entre documentos²²⁴, interconexión entre la documentación normativa, judicial y bibliográfica, y actualización permanente. Las bases de datos jurídicas responden al modelo de recuperación tradicional asentado en el uso de un fichero tradicional, un fichero inverso y un lenguaje de recuperación basado en el álgebra de Boole; si bien editoriales como La Ley o Aranzadi han comenzado a introducir la presentación de resultados ordenados por relevancia²²⁵. Por tanto la interacción entre usuario y sistema se efectúa mediante la realización a la base de datos de preguntas cuyos términos han de coincidir exactamente con los que constituyen los puntos de acceso de sus registros.

En el caso de las bases de datos legislativas, los sistemas de recuperación han de adaptarse adecuadamente a las peculiaridades y complejidad de los documentos normativos. Ello requiere la estructuración de los registros en una serie de campos y subcampos que se acomoden a la tipología concreta tratada. La unidad documental es la norma. Además del texto completo de la disposición, un tratamiento documental previo asegura la recuperación de la información y garantiza la seguridad jurídica, mediante la normalización conceptual y la resolución de las interconexiones entre las normas.

²²³ Dado que la información parlamentaria, legislativa y jurisprudencial ha de ser almacenada de forma íntegra para responder a las necesidades reales de los usuarios.

²²⁴ Relaciones normativas entre documentos anteriores y posteriores que complementan, amplían, modifican e interpretan una disposición.

²²⁵ ALVITE DÍEZ, M^a. Las herramientas terminológicas en los sistemas de información jurídica. *Scire*, vol.10, n°1, 2004, pp. 77-90.

En los sistemas jurisprudenciales la unidad documental es la resolución. Además del texto completo, con las particularidades adoptadas por cada sistema, los registros incorporan el análisis documental, dirigido fundamentalmente a la asignación de descriptores y al tratamiento de las conexiones con la normativa y la jurisprudencia tratada en la correspondiente sentencia. Las bases de datos habrán de adaptarse a las características de la documentación judicial, articulando los registros en una serie de campos y subcampos ajustados a la correspondiente tipología del documento.

Los sistemas tienen en cuenta la naturaleza de la información procesada y tratan de adaptarse a las necesidades del usuario. Si bien la información jurídica forma parte de la vida cotidiana o profesional de cualquier ciudadano, las bases de datos evidencian un tratamiento y diseño orientado en gran medida a los especialistas en Derecho. En este sentido, tal vez Iberlex sea la más alejada del estándar jurídico. El estudio sobre el comportamiento de los usuarios especialistas durante el proceso de búsqueda señala una clara tendencia de los usuarios a concentrarse en aprender lo mínimamente imprescindible sobre el sistema para obtener resultados relevantes, y seguidamente a concentrar su esfuerzo en su lectura y valoración, resultados que coinciden con otros trabajos de evaluación de interfaces. Se constata además la estrecha relación que existe entre los sistemas y las necesidades, habilidades y métodos de trabajo de los profesionales del Derecho. Las interfaces de las bases de datos jurídicas confirman esto. En la recuperación de información legislativa se produce una combinación de información conceptual e información factual, que los sistemas resuelven con la introducción de campos de búsqueda suficientes para dar respuesta adecuada a consultas por fechas, rango, número o fecha del boletín, título, texto, voces o índice. Los sistemas jurisprudenciales pretenden la mayor integración posible con los hábitos de trabajo y la práctica jurídica cotidiana de los especialistas del Derecho y atienden a la naturaleza de la información tratada, sea ésta factual o conceptual. Existe una gran coincidencia en los campos de consulta proporcionados por los sistemas: resolución, jurisdicción, ponente, disposiciones estudiadas, texto y voces.

Los usuarios a los que se dirigen estos sistemas cifran sus expectativas en lograr la máxima expectación en las respuestas, y ésta se alcanza cuando la interrogación se realiza sobre campos factuales: fechas de emisión, de publicación o número de disposición. Se trata de búsquedas con una elevada tasa de precisión sobre las que se pueden emitir juicios de

relevancia a priori, que no suelen provocar problemas de recuperación y que están disponibles a través de muchos sistemas gratuitos disponibles en Internet, si bien es cierto que con una menor cobertura temporal. Sin minusvalorar la eficacia de estas consultas, creemos que la capacidad real de estos sistemas se determina por su capacidad para recuperar la información temática o conceptual tratada. Los puntos de acceso por materias son fundamentales y su estudio es central en el diseño de sistemas de información; si bien es cierto que no podemos constreñir la recuperación temática a la proporcionada por los puntos de acceso de materias despreciando otros puntos de acceso descriptivos. En el caso de las bases de datos jurisprudenciales, estos puntos de acceso suponen una ayuda inestimable para conocer el articulado estudiado y la línea jurisprudencial seguida en una resolución. Tampoco se debe soslayar la trascendencia de los enlaces hipertextuales en la recuperación de información, en cuyo uso fueron pioneros estos sistemas.

Recuperación de información y organización del conocimiento son conceptos imbricados de modo indisoluble. Mooers²²⁶ acuñó el término de “lenguajes de recuperación de información”, para quien²²⁷:

La recuperación de información es la denominación que se le da al proceso o método por el cual un potencial usuario es capaz de convertir sus demandas informativas en una lista actualizada de citas a documentos en una colección con información útil para él. La recuperación de información incluye tanto los aspectos intelectuales para la descripción de la información como aquellos necesarios para la búsqueda e incluso cualquier sistema, técnica o máquina que pueda ser empleada para la operación de recuperación.

Con ello se refiere, de modo genérico, a aquellos sistemas artificiales de signos normalizados que hacen posible la representación del contenido de los documentos, para facilitar la recuperación de información. La recuperación del texto completo de los documentos es el objetivo final y es, a su vez, la idea que impulsa el desarrollo de los puntos de acceso por materias. Así, en los sistemas de recuperación de información jurídica, cada palabra o combinación de palabras se convierten en potenciales puntos de

²²⁶ MOOERS, CALVIN N. Zatocoding applied to mechanical organization of knowledge. *American Documentation*, nº2, 1951, pp.20-32.

²²⁷ SARACEVIC, T. Evaluation of evaluation in information retrieval. En *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*. New York : ACM, 1995, p.139. En: <http://delivery.acm.org/>. Consultado 15/12/2009.

acceso temáticos. Esto nos lleva a destacar la importancia de incluir algún tipo de análisis intelectual en campos de valor añadido que complementen las búsquedas efectuadas sobre el texto completo. Sin embargo, el principal inconveniente de las búsquedas en lenguaje jurídico sobre bases de datos a texto completo es la inconsistencia en el uso de las palabras y formas propias del lenguaje, lo que nos debe llevar a reclamar más investigaciones que demuestren por un lado la eficacia de la indización en lenguaje controlado y por otro, determinar cuál es el tipo de lenguaje más apropiado para cada disciplina.

Rowley declaró que los lenguajes natural y controlado deben ser usados conjuntamente²²⁸. Hjørland defiende la necesidad en centros de documentación especializada de profesionales de la información, argumentando que los documentos en sí mismos pueden no incluir implícitamente la materia de la que tratan o bien hacerlo de modo erróneo. Su trabajo sobre organización del conocimiento aglutina líneas de investigación como el desarrollo del procesamiento del lenguaje natural, la indización automática, el uso del hipertexto, las evaluaciones sobre el rendimiento de sistemas, y una tendencia que creemos afecta de manera específica a la documentación jurídica, el valor de los términos que aparecen en el texto según la zona del mismo en la que aparecen; y precisamente, es éste uno de los aspectos explotados en los proyectos de aplicación de XML al tratamiento y recuperación semántica de la información jurídica²²⁹.

El campo jurídico ha sido pionero en la recuperación en línea a texto completo. El primer sistema aplicado a la recuperación de una gran colección legislativa se presentó en el año 1960 por el Health Law Center de la universidad de Pittsburg. Utilizaba ya un tesoro de asistencia en el proceso de búsqueda. Lancaster se refiere a este tesoro como una simple compilación de palabras con significados similares, con una estructura semejante a la del Roget's Thesaurus que suponía una importante ayuda para la búsqueda al evitar que el usuario tuviera que pensar en todas las posibles palabras para expresar una idea concreta. Se trataba de un tesoro simple que²³⁰: *Carecía de un significativo grado de "estructura"*.

²²⁸ ROWLEY, J. *Organizing Knowledge: an introduction to managing access to information*. Hampshire : Ashgate, 2008, p.167.

²²⁹ HJORLAND, B. Fundamentals of knowledge organization. En *Actas del IV Coloquio Internacional de Ciencias de la Documentación y VI Congreso del Capítulo Español de ISKO*. Salamanca : Universidad de Salamanca, 2003, p.83-116.

²³⁰ LANCASTER, F.W. *El control del vocabulario en la recuperación de la información*. Valencia : Universitat de Valencia, 2002, p.215ss.

Desde entonces han sido numerosos los experimentos evaluativos a los que han sido sometidos los diversos sistemas de recuperación legal, hallándose conclusiones dispares sobre su rendimiento en términos de exhaustividad y precisión; sin embargo, todos ellos se centran en el esfuerzo que el usuario necesita realizar durante la consulta en texto libre para predecir las palabras y frases necesarias para poder lograr la recuperación de documentos relevantes. Para intuir la complejidad que supone la predicción de los términos de consulta, se ha de tener presente que, en el lenguaje jurídico, atendiendo a la procedencia terminológica, podemos encontrar tanto vocabulario estrictamente jurídico, como lenguaje técnico o especializado de otras especialidades de las ciencias sociales. El lenguaje controlado elaborado para la indización y recuperación de la documentación jurídica se enfrenta a la dificultad de atender a las tres parcelas puesto que no cabe la expresión de una ley o sentencia en puro lenguaje de técnica jurídica²³¹.

En nuestro país las bases de datos jurídicas, tal vez por herencia de los repositorios impresos en los que algunas de ellas tienen su origen, cuentan con instrumentos de control terminológico, en muchos casos asociados a un campo que recibe la denominación de “Voz”. Ya los repertorios decimonónicos recurrían a índices de voces para organizar el complejo lenguaje tratado en la documentación jurídica, esto es, listas autorizadas de términos preferentes que hacen referencia a un marginal que conduce al documento²³². Creemos que los sistemas actuales intentan precisamente adaptarse, por un lado, al modo de trabajo de los profesionales jurídicos apegados a las listas autorizadas de entradas; y, por el otro, evitar las dificultades de la búsqueda en lenguaje natural sobre documentos de la extensión que alcanzan muchos de los registros normativos o jurisprudenciales almacenados en este tipo de bases de datos. El usuario ha de esmerarse para predecir inequívocamente los términos de búsqueda y para formular una ecuación de consulta efectiva que evite que el sistema devuelva un elevado número de documentos no relevantes.

Hjörland se refiere a diferentes tipos de puntos de acceso temáticos en cuanto que describen de modos diversos la materia dada de un documento: con mayor o menor exhaustividad, de forma más o menos general o específica, de modo abierto o cerrado,

²³¹ ALVITE DÍEZ, M^a. Tendencias de la investigación sobre recuperación de información jurídica. *Revista Española de Documentación Científica*, vol.26, nº2, 2003, p.199-200.

²³² Id.

pensando en el valor futuro dentro de la base de datos²³³. Convenimos con estos autores en conceder el máximo valor a aquellos puntos de acceso temático que hacen posible al usuario identificar los documentos de superior relevancia, esto es, los que en la base de datos hacen más visibles a los documentos más relevantes en detrimento de los menos relevantes. Los sistemas jurídicos que nos ocupan disponen de varios campos para acceder a información temática y, sin embargo, todos ellos establecen en su interfaz un campo distintivo que sirve de punto de acceso al contenido mediante el uso de vocabulario controlado. Cuando se pretende recuperar información cognitiva, evitar el ruido, simplificar las ecuaciones de búsqueda y contextualizar los términos, estas bases de datos precisan la utilización de lenguaje controlado.

Nos sumamos al amplio consenso de la literatura científica acerca de la complementariedad entre el lenguaje libre y el controlado. López Alonso²³⁴ presenta una recapitulación de esta controversia sobre la que Gil Urdiciain, tras una comparación del rendimiento de diferentes tesauros españoles frente al lenguaje natural en la recuperación de información, concluye señalando que el control del vocabulario es un factor determinante del éxito en el proceso de recuperación, dado que el lenguaje controlado neutraliza las deficiencias del lenguaje libre²³⁵.

Con la irrupción de Internet y la digitalización, las organizaciones e instituciones se vieron sometidas a un importante proceso de cambio, afrontado muchas veces a partir de la inexperiencia con respecto a las nuevas tecnologías. Aquellas instituciones, tales como el BOE, trabajaban directamente con procesos de gestión, almacenamiento y recuperación de información casi exclusivamente en papel, con el empleo de unas pocas bases de datos jurídicas en determinados centros de apoyo a la investigación como el CSIC.

Esta necesidad de integrar las nuevas tecnologías de la información en los procesos de gestión y recuperación de información se vivieron de manera particularmente intensa en las bibliotecas, archivos y centros de documentación jurídica. Espacios complejos de

²³³ HJORLAND, B. *Information seeking and subject representation: an activity-theoretical approach to information science*. Westport : Greenwood Press, 1997, p.123ss.

²³⁴ LOPEZ ALONSO, M.A. Evolución histórica y tendencias observables en los tesauros. *Boletín Millares Carlo*, nº17, 1998, p.360ss.

²³⁵ GIL URDICIAIN, B. Evaluación del rendimiento de tesauros españoles en sistemas de recuperación de información. *Revista española de documentación científica*, v.21, nº3, 1998, p.286ss.

almacenamiento y distribución de colecciones documentales difícilmente afrontables. En este nuevo contexto digital no solo se trataba de diseñar sistemas automatizados de bibliotecas, sino también de reproducir obras completas facsimilares y, sobre todo, de decidir qué nuevos canales de comunicación y servicio se abrirían hacia los usuarios. En muchos de los casos las instituciones se encontraron en la tesitura obligada de dar salida externa a una considerable cantidad de información a través del acceso digital. El desafío era múltiple, adaptar la cultura corporativa de sus empleados, renovar el hardware y software de manera fiable y afrontar el hecho de que, a partir de ese momento, su presencia exterior expondría los catálogos no sólo a los usuarios que físicamente acuden a sus centros sino a cualquier usuario del mundo con acceso a Internet.

Este desafío ha sido afrontado mediante soluciones bastante heterogéneas desde el punto de vista tecnológico y editorial. Diferentes equipos de profesionales se han enfrentado a la digitalización de grandes fuentes de documentación con resultados dispares y con soluciones formuladas a partir de la experimentación con sus propios documentos.

Buena parte de los retos han provenido de las grandes instituciones públicas como el CSIC, la Biblioteca Nacional, o las distintas redes de bibliotecas, bien las públicas del Estado vertebradas a través de la red REBECA, o bien las universitarias organizadas en torno a consorcios como MADROÑO, CBUC o REBIUN, o bien las dependientes de un organismo de investigación como el CSIC. Todas estas bibliotecas eran depositarias de fondos heterogéneos susceptibles de ser puestos a disposición del público a través de Internet, pero con la dificultad de tratar de manera homogénea unas colecciones constituidas por unidades documentales en muy diversos formatos, cada uno de ellos con un tratamiento específico. Asimismo, ya no se trata de garantizar el acceso óptimo a los investigadores a través de las redes sino de abrir nuevos canales de exhibición para nuevos públicos. De esta forma, el British Museum incrementó la circulación sobre algunos de sus fondos a través de una base de datos, Compass²³⁶, sobre la que trazan rutas electrónicas. La biblioteca nacional de Francia mantiene una estrategia similar a través de su corpus Gallica²³⁷, una selección digitalizada de sus fondos sobre la que se establecen rutas de navegación. La New Cork Public Library ha dado un paso más allá organizando sus fondos multimedia relativos a las migraciones del colectivo de raza negra dentro de los Estados

²³⁶ En: www.thebritishmuseum.ac.uk/compass. Consultado el 20/12/2009

²³⁷ En: <http://gallica.bnf.fr>. Consultado el 20/12/2009

Unidos en un amplio contenedor documental, In Motion²³⁸, destinado al público en general y específicamente a estudiantes para quienes se han elaborado materiales didácticos.

En los últimos años, muchos sitios han aglutinado varias bases de datos que distribuyen información escalada a grandes audiencias, reuniendo en su interior amplias combinaciones de bases de datos. Un ejemplo de ello es la Biblioteca Miguel de Cervantes, que mantiene un conjunto heterogéneo de materiales relativos a la cultura hispánica, y que han ido evolucionando desde un núcleo original de literatura para convertirse en un contenedor de diferentes documentos y bases de datos históricas, científicas y sociales con un formato sensiblemente diferenciado en función de su contenido.

Las soluciones fueron en gran medida, ad hoc, basadas en experiencias empíricas, y destinadas a resolver problemas en el análisis formal de los registros bibliográficos. En muchos archivos bibliográficos, sin embargo, la presencia web no había sido pensada para un usuario final, interesado en el acceso inmediato a la información.

En otras áreas de la red el proceso era opuesto. Durante muchos años los diseñadores se centraron en preparar envoltorios impactantes avanzando en sus investigaciones empíricas sobre el tratamiento del color en la web, el empleo de tecnologías de formato de textos como CSS o la relevancia del uso de la cuarta dimensión de las páginas, el movimiento y el multimedia a través de las tecnologías DHTML, Shockwave y Flash. Buena parte de ese gran número de proyectos web que se estaban generando en esos primeros años partían de una limitadísima cantidad de contenidos y no estaban pensados para soportar un continuo incremento de audiencia y materiales publicados en los siguientes años. De esta forma, la experimentación en formatos vivió durante un tiempo separada de la propia gestión del contenido.

Los vocabularios empleados durante la navegación y búsqueda de información en la web son producto de las diferentes arquitecturas de información empleadas en los distintos sistemas de información²³⁹. La relación entre la arquitectura en los distintos sistemas de información y los vocabularios de búsqueda y navegación tiene una larga trayectoria en la

²³⁸ En: <http://www.inmotionaame.org>. Consultado el 20/12/2009

²³⁹ WARNER, A. Information architecture and vocabularies for browse and search. En GILCHRIST, A. ; MAHON, B. *Information Architecture. Designing information environments for purpose*. London : Facet, 2004, p. 177-178.

historia de la Documentación y las bibliotecas, pero para el propósito de nuestra tesis nos interesa discutir la aplicación de los vocabularios de búsqueda a la arquitectura de información.

Hasta ahora, la investigación en lenguajes controlados se ha centrado en sus propiedades, características de diseño y dominios de aplicación. Los dos principales tipos de vocabularios son aquellos orientados a la navegación como es el caso de las taxonomías; y aquellos que se orientan a las búsquedas, como es el caso de los tesauros. Sin embargo existen cuatro opciones tecnológico-documentales para su aplicación a la búsqueda en la web:

Taxonomías que no pueden combinar etiquetas para la navegación. Se trata de navegación sin facetas y la navegación más antigua en la web. Supone un conjunto discreto de archivos, muchos de los cuales con hiperenlaces a otros recursos de información. La dificultad para los diseñadores de estos sistemas reside en determinar qué hiperenlaces seleccionar a objetos de información en la Web. Esta opción es adoptada fundamentalmente por organizaciones con sitios web poco densos (con escasos recursos de información). Es tecnológicamente la opción más simple pero no necesita de una arquitectura de información. Esto supone que a medida que ha ido creciendo el sitio, el control de sus recursos se ha vuelto incontrolable. Esta situación se ha ido agravando en las instituciones jurídicas que han pasado de un sistema centralizado a uno descentralizado a medida que se han ido transfiriendo las competencias a las Comunidades Autónomas. Esto ha generado numerosos sitios web de la administración con hiperenlaces rotos y con falta de consistencia en el diseño del sitio web entre los propósitos para los que fueron diseñados y la comunidad de usuarios a la que se orientan. No obstante, las organizaciones que no han crecido y se han mantenido centralizadas en el diseño y mantenimiento de su sitio web no se han visto afectadas y disponen de un sitio web consolidado.

Navegación por taxonomías, que pueden combinar etiquetas en el interfaz. Esta segunda opción crea una estructura de navegación que proporciona una combinación de etiquetas en el interfaz. Para ello se crea una base de datos central en la cual los registros son etiquetados con las etiquetas de cada una de las facetas de una taxonomía. La base de datos contiene estas etiquetas junto con una lista de URL's que apuntan a los objetos de información de la Web. La ventaja de este sistema es que los cambios se efectúan a través

de una base de datos centralizada más que por medio de una colección descentralizada de archivos.

Un buen ejemplo de arquitectura de información con este tipo de vocabulario es el empleados en el diseño de sitio del Centro de Estudios Políticos y Constitucionales (CEPC)²⁴⁰. Las distintas secciones del centro son etiquetadas por sus funciones, competencias y secciones según el ámbito del que se trate. Los usuarios, en vez de buscar en el motor de búsqueda del sitio por algún servicio del centro pueden seleccionar y especificar por medio de la navegación aquellos elementos que les resulten de más interés para la navegación por el sitio por medio de un proceso de especificación de etiquetas para las búsquedas.

Lo fundamental de este interfaz es un sistema que efectúa una búsqueda y la pasa a una base de datos. La ventaja de este tipo de sistema es su usabilidad para la comunidad de usuarios del sitio web al permitirles un acceso intuitivo a las distintas páginas web que constituyen el sitio. La desventaja es que sólo funciona bien para sistemas que no son demasiado complejos, especialmente para sitios con mucha información textual y no demasiados recursos multimedia. Sin embargo a medida que crece la colección y ésta se hace más compleja tanto por las relaciones entre los documentos como de éstos con sus correspondientes etiquetas, el número de facetas y etiquetas empleadas se vuelve demasiado numerosa con el consiguiente peligro de que los usuarios lleguen a perderse.

²⁴⁰ CENTRO DE ESTUDIOS POLÍTICOS Y CONSTITUCIONALES. CEPC. En: <http://www.cepc.es/>. Consultado el 15/12/2009.

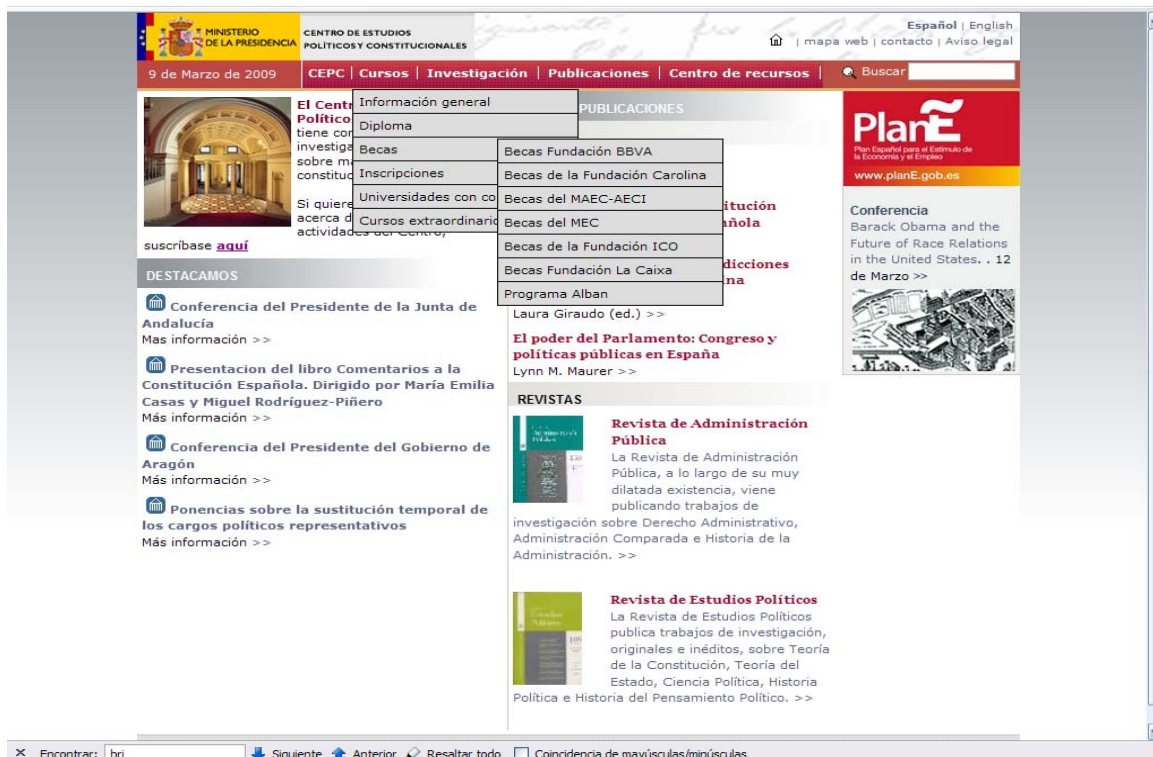


Figura 10. Ejemplo de sitio web con búsqueda por etiquetas. El sitio web del CEPC.

Sistemas de búsqueda que no pueden combinar descriptores para la búsqueda. La tercera opción se refiere fundamentalmente a los sitios en que los algoritmos de indización estadística se aplican a registros desestructurados. Este es el segundo vocabulario de control más antiguo en la web y es empleado por la mayoría de los motores de búsqueda comerciales que hay en la web. Muchas organizaciones utilizan esta opción para sus intranets dado que no necesita contar ni con un software caro ni con una base de datos que estructure la información.

Sistemas de búsqueda que pueden combinar descriptores para la búsqueda. La opción final es la más reciente, implica una sofisticación técnica mayor y por tanto es más costosa desde el punto de vista económico dado que la organización debe invertir en un sistema de gestión de contenidos para alojar metadatos, algún tipo de infraestructura de etiquetado y un motor de búsqueda que se encargará de la naturaleza estructurada de sus datos. Además las organizaciones deben ser capaces de confiar en la capacidad de sus usuarios para manejar el sistema para que puedan beneficiarse de su potencial. La ventaja que tiene este sistema es que a través de esta estructura y el etiquetado de alta calidad, las búsquedas están más orientadas a producir mejores resultados. A causa del tiempo y coste que tiene cargar estos sistemas sólo se emplean en grandes organizaciones.

La discusión se centra sobre el rango de opciones para vocabularios controlados, sus ventajas y desventajas, y cómo estos pueden adecuarse a una arquitectura básica de información. Son muchas las variables y opciones a tener en cuenta tales como el número de recursos, el dominio de la colección, la disponibilidad de la tecnología necesaria y el nivel de formación digital de nuestra comunidad de usuarios. En cualquier caso, en arquitectura de información no existen leyes acerca de los vocabularios controlados a emplear, pero sí podemos apuntar algunas tendencias. Una de ellas es que la navegación y búsqueda satisfacen necesidades de información diferentes y por ello se cuenta con ellas cada vez en más sitios, pero especialmente en aquellos que son muy densos.

Otra tendencia que se está produciendo es la creación de sistemas híbridos, especialmente con sitios web muy densos (con muchas páginas web). Implica entender el conjunto de recursos informativos de la colección como etiquetada en términos de sus características y su valor en la colección. Así, por ejemplo, se hace una descripción más exhaustiva de aquellos documentos con más valor para la misión de la organización, y se hace una descripción más superficial para los documentos menos pertinentes.

Cualquier organización debe ser capaz de evaluar la utilidad del diseño de su arquitectura de información y la aplicación de vocabularios controlados y taxonomías frente a las demandas de su comunidad de usuarios. Esta evaluación depende de múltiples factores tales como el contexto de actividad de la organización, características intrínsecas del sistema de información como la densidad del sitio, la tecnología sobre la que se sostiene el sitio, así como las habilidades informativas de los miembros de la comunidad a la que se destina el sitio web y su sistema de información. Las respuestas a las preguntas complejas de los vocabularios controlados difieren para cada organización ya que cada una es única en sus recursos y orientación a sus usuarios.

3.3. El lenguaje de especialidad. El español jurídico.

Una variedad de la lengua es la profesional y académica²⁴¹, lo que ha dado pie a una nueva línea de investigación denominada el “español profesional y académico”, también conocida con las siglas EPA. Forma parte de lo que se denomina “lenguas de especialidad”, término

²⁴¹ ALCARAZ VARÓ, E. ; HUGHES, B. *El Español Jurídico*. Barcelona : Ariel, 2002. p. 15.

tomado del francés *langues de spécialité*²⁴², y que alude al lenguaje específico que utilizan los profesionales y expertos para transmitir información y, cuando es preciso, para refinar los términos, los conceptos y los saberes de una determinada área de conocimiento, confirmando los ya existentes, matizando el ámbito de su aplicación o modificándolos total o parcialmente²⁴³.

Con anterioridad, Saussure ya las había mencionado con el nombre de “lenguas especiales”; en su opinión, estas lenguas “son fomentadas por un avanzado grado de civilización” y, entre ellas, cita la *lingua jurídica*²⁴⁴.

En la tradición lingüística inglesa estas lenguas de especialidad se llaman “registros”, los cuales se definen como las variedades de una determinada lengua destinadas a cumplir un fin comunicativo en un marco profesional o académico concreto²⁴⁵. Cada una de estas variedades es, en principio, un subconjunto de rasgos léxicos y gramaticales que, por repetirse con frecuencia y regularidad, le confieren cierto aire de unidad diferenciada.

El “español profesional y académico”, en cada una de sus variedades, es una lengua *profesional* porque es la que emplean los profesionales en su comunicación diaria, y es también *académica* porque antes de haber sido utilizada en cada ambiente profesional, fue enseñada y aprendida en la universidad. En esta institución se perciben dos movimientos epistemológicos complementarios, con la implicación lingüística que entraña toda iniciativa de esta naturaleza: el flujo de información hacia las profesiones y el reflujo proveniente de ellas. De esta forma, la universidad da conocimientos e información lingüística, pero también recibe, renovándose y perfeccionándose por el continuo contacto con la realidad que le ofrece este movimiento circular.

Una de las variantes más destacadas del EPA es el español jurídico, ya que las relaciones sociales están regidas por el orden jurídico. El español, por el peso que posee en la comunidad internacional, tiene mucho que decir en esta vertiente profesional y académica.

²⁴² Id.

²⁴³ Id.

²⁴⁴ SAUSSURE, F. *Curso de Lingüística General*. Buenos Aires : Losada, 2002. p. 54-55.

²⁴⁵ HALLIDAY, M.A.K. *El lenguaje como semiótica social: la interpretación social del lenguaje y del significado*. México : Fondo de Cultura Económica, 2001, p. 34.

En el análisis del español jurídico podemos distinguir, desde un punto de vista lingüístico, tres objetos de estudio²⁴⁶:

1. Un vocabulario muy singular, que constituye el núcleo de este lenguaje especializado.
2. Unas tendencias sintácticas y estilísticas muy idiosincrásicas.
3. Unos géneros profesionales propios e inconfundibles, como la ley, la sentencia, el contrato, el conocimiento de embarque, el crédito documentario, etc., que sirven para marcar bien los límites de estas dos variantes del español profesional y académico respecto de las demás.

El español jurídico está poco estudiado, ni siquiera existe acuerdo sobre si se debe considerar o no una lengua de especialidad. Para algunos autores su denominación sería “lenguaje técnico-jurídico” pero prefieren el de lenguas especializadas de la Jurisprudencia y la Legislación, puesto que se trata de “lenguas” que son plenamente “español”, aunque al mismo tiempo son el instrumento transmisor de saberes y de prácticas profesionales²⁴⁷.

Cuando se habla del español jurídico es porque se estima²⁴⁸ que el Lenguaje del Derecho y no el lenguaje común especial tiene conciencia de lenguaje autónomo por su vinculación a un área de saber concreta que es la materia jurídica y en la cual toda su significación es jurídica y por tanto merece salir del encierro del lenguaje vulgar para alcanzar un modo de expresión particular. Sin embargo, no puede pretenderse que el lenguaje de una ciencia se independice del común ya que ello representaría la creación de otro idioma necesitado de traducción y de unos descriptores propios para su recuperación. El nombre de común para el lenguaje base significa su presencia en todos los lenguajes del mismo idioma, porque es el que presta el material lingüístico básico y sus reglas gramaticales y fonéticas. Pero su mantenimiento como lenguaje común no requiere privar de un tratamiento especial a aquellos lenguajes con una significación ligada a una Ciencia, como es el Derecho. En el caso de este lenguaje ocurre que da la impresión de que en muy poco difiere del lenguaje natural, por ser escasos los términos que le separan del lenguaje ordinario tal y como ocurre en el lenguaje de las Ciencias Naturales. No obstante, está tan ligado el Derecho a la vida

²⁴⁶ BROUGHTON, V. Op. Cit. p. 16.

²⁴⁷ Id.

²⁴⁸ BURGO y MARCHÁN, Ángel Martín del. *El lenguaje del derecho*. Barcelona: Bosch, 2000. pp.129-130.

del hombre, como para que su lenguaje, aunque jurídico, esté compuesto en su mayoría de palabras de empleo ambivalente con coexistencia tanto en la vida jurídica como en la ordinaria.

Al hablar de la entidad del español jurídico se afirma que las distintas áreas del conocimiento cuentan, si no con un lenguaje propio, sí con particularidades y modismos semánticos, y encierran en su fondo cierto artificio por cuanto que suponen un apartamiento del uso común; y es cierto que en torno al Derecho se ha ido formando un lenguaje técnico especializado, que si es familiar para sus cultivadores sorprende a los profanos.

El español jurídico no constituye un bloque monolítico. Se pueden distinguir al menos cuatro variantes: el legislativo o de los textos legales, el jurisdiccional o de los jueces (sentencias, autos, providencias, etc.), el administrativo o de la Administración del Estado (instancias, expedientes, etc.), y el notarial (testamentos, escrituras de compraventa, poderes, etc.). Cada uno de ellos posee rasgos peculiares, pero el que está por encima de todos los demás es el español de las leyes, también llamado el jurídico-normativo, que marca las pautas que habrán de seguir el lenguaje de los jueces, el de la Administración y el de los notarios. La ley está por encima de los demás géneros jurídicos porque es la norma jurídica suprema de carácter general e impersonal, aplicable a todos, elaborada y aprobada por las Cortes Generales; si la ley es el género profesional por excelencia del Poder Legislativo, la sentencia lo es del Poder Judicial.

Las sombras del español jurídico nacen de su opacidad y de su falta de naturalidad. La opacidad conduce irremediabilmente al “oscurantismo” que, de acuerdo con la única acepción que de este término dan los diccionarios, es la oposición sistemática a que se difunda la instrucción entre las clases populares. Contribuyen a esa opacidad no sólo los significados de términos especializados sino también y particularmente muchos usos lingüísticos incorrectos, por ejemplo, los frecuentes anacolutos y el abuso del gerundio; dado que, en el “gerundio jurídico” reside gran parte de la ininteligibilidad de los artículos legales y de las sentencias, ininteligibilidad que se acentúa cuando se tiende a su acumulación en una misma frase.

La “naturalidad” en la expresión está ausente en bastantes sentencias, en las que se pueden encontrar expresiones como “La intencionalidad de causación del resultado”, o “Ante lo que el hecho de la hipotética reversibilidad de esa situación dice tanto como lo que ...”, que podrían haberse redactado con mayor claridad y sencillez.

3.3.1 El español jurídico en la informática jurídica documental.

Ya se han hecho trabajos de investigación sobre la naturaleza y las características del español jurídico²⁴⁹. Sin embargo, el campo es muy extenso, y donde parece que se agranda aún más es en el campo de la Documentación. Se abren dos grandes líneas de investigación:

I) Oracional. En la perspectiva oracional, el vocabulario, la morfosintaxis y la estilística ofrecen amplias posibilidades para la mejora de la recuperación de información en bases de datos de Derecho.

Una de ellas es la Lexicología²⁵⁰. Es probablemente la especialidad más abierta al estudio del español jurídico, ya que, como ciencia que se ocupa de describir las palabras de la lengua, su finalidad es explicar de la forma más adecuada posible el funcionamiento léxico del jurista o del legislador. Por tanto son muchas las áreas de investigación que se ocupan de ella, siendo la primera de ellas la traducción. Sin embargo para la informática jurídica documental tiene una presencia fundamental en relación con la indización y recuperación en bases de datos o bien para la clasificación y ordenación multilingüe de los documentos jurídicos en instrumentos documentales, como Tesoros, elaborados por Instituciones Internacionales como es el caso del Tesoro EUROVOC de la Unión Europea²⁵¹. Creado en 1984 por la Oficina de Publicaciones Oficiales de la Unión Europea, con el objeto de poder indizar todas las versiones lingüísticas de las disposiciones emanadas del Parlamento Europeo de una sola vez, pero que luego sería incorporado por los servicios de documentación y archivo de los distintos parlamentos nacionales de cada estado integrante de la Unión Europea y de países europeos extracomunitarios.

En el español jurídico se distinguen tres clases de términos:

²⁴⁹ Es el caso de la obra de Enrique Alcaraz Varó, o de Ángel Martín del Burgo y Marchán, o de Jesús Prieto de Pedro entre otros.

²⁵⁰ CABRÉ, Mª Teresa. *La Terminología: Teoría, metodología, aplicaciones*. Barcelona: Antártida/Empúries, 1993. pp.80-82.

²⁵¹ CUETO APARICIO, M. El nuevo tesoro EUROVOC como instrumento de recuperación multilingüe. En *II Jornadas de Tratamiento y Recuperación de la Información*. Leganés : Universidad Carlos III, 2003.

Técnicos: aquellos que sólo se utilizan en el español jurídico, aunque alguno de ellos haya pasado al lenguaje común.

Semitécnicos: aquellos que, sin perder el significado que tienen con el lenguaje común, han adquirido otro en el de especialidad.

Generales: de uso frecuente en una especialidad, pero es el más amplio.

Pero dentro de la lexicología hay muchas otras líneas de investigación. Están por hacer, entre otros, los siguientes estudios:

a) El de las combinaciones léxicas de las principales palabras del español jurídico; por ejemplo, las “leyes” autorizan, derogan, invalidan, prescriben, revocan, etc.

b) El de las sinonimias parciales de los términos jurídicos como, por ejemplo, los sinónimos de “disponer”: determinar, establecer, fijar, regular, precisar, etc.

c) El de los campos semánticos también es un área interesante; la comprensión de un término es más completa cuando se le encuadra en su campo semántico; por ejemplo, en el de la “demanda” encontraríamos “demandante”, “demandado”, “actor”, “pretensión”, “recurso”, etc.

d) El de los “falsos amigos” jurídico-económicos, que siempre ha atraído a varios investigadores: por ejemplo “demandante” es *demandeur* en francés, pero “demandado” no es *demandé* sino *défendeur*, ya que la palabra francesa *demandé* significa “persona que ha sido llamada al teléfono o por teléfono”, y *défenseur* es el “abogado defensor”; igualmente, “prorrogar” no es *proroge* en inglés sino *extend*, puesto que *proroge* equivale a “suspender”, “aplazar” o “interrumpir”.

II) Supraoracional. En el campo de la lingüística supraoracional, dos líneas de investigación muy fecundas son las que corresponden a los tipos textuales, también llamados géneros profesionales, tanto los orales como los escritos, (la ley, la sentencia, la entrevista, el contrato, el conocimiento de embarque, etc.) y a las modalidades discursivas (narración, descripción, exposición, etc.). El estudio de las macroestructuras de los géneros

profesionales, de sus estrategias comunicativas ofrece muchos problemas a los investigadores.

3.3.2. Herramientas lingüísticas para el análisis del lenguaje jurídico.

3.3.2.1 El proyecto Wordnet.

Wordnet es un sistema de referencia léxico en línea cuyo diseño se inspira en las teorías psicolingüísticas de la memoria léxica del hombre, de hecho es un proyecto desarrollado en el Laboratorio de Ciencias Cognitivas de la Universidad de Princeton²⁵². Su objetivo es convertirse en una base de datos léxico-conceptual constituida tanto por unidades léxicas como por las relaciones entre ellas y estructurado en forma de red semántica relacional²⁵³. Para ello se busca que sea capaz de unir las características de los diccionarios electrónicos con las características de los Tesoros online. Los primeros resultados de esta investigación dieron sus frutos con la primera versión online de Wordnet en 1993. Diez años más tarde Wordnet²⁵⁴ ha alcanzado la versión 1.7.

²⁵² FELLBAUM, Christiane (ed.). *Wordnet: An Electronical LexicalDatabase*. Cambridge: MIT, 1998.

²⁵³ MORATO, J; MARZAL, M.A.; LLORÉNS, J; MOREIRO, J. WordNet applications. En: SOJKA, P.; PALA, K. ; SMRZ, P. ; FELLBAUM, CH. ; VOSSEN, P. *GWC 2004: Proceedings of the Second International WordNet Conference*. Brno : Masaryk University, 2004. pp. 270-278.

²⁵⁴ WORDNET. En: www.cogsci.princeton.edu Consultado el 16/12/2009.

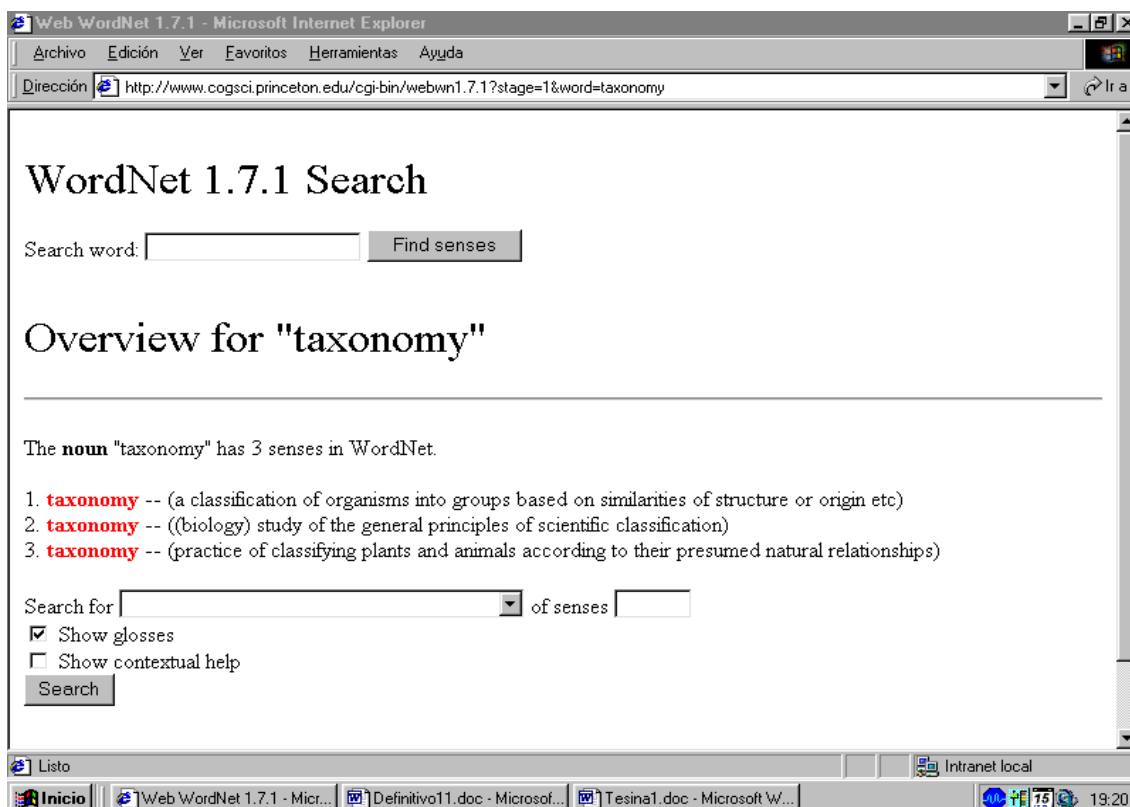


Figura 11. Resultados de una búsqueda con Wordnet 1.7.1.

El contenido de Wordnet está basado en que la unidad básica es la palabra, por lo tanto no se descomponen las palabras en unidades de significado menor, aunque una comparación entre componentes analíticos de las palabras revela algunos aspectos comunes. Por ejemplo, la relación Causa que enlaza muchos pares de verbos es considerado un primitivo semántico en el análisis que descompone significados en términos de estructuras conceptuales léxicas. Los sustantivos, verbos, adjetivos y adverbios ingleses están organizados en colecciones de sinónimos, cada una representando un concepto léxico destacable. Distintos tipos de relaciones enlazan los grupos sinonímicos.

El objetivo de Wordnet es la construcción de una base de datos léxica multilingüe con wordnets. Esta base de datos, con información acerca de sustantivos, verbos, adjetivos y adverbios en inglés, está organizada en torno a la noción de synset. Un synset es una colección de palabras con la misma raíz que puede ser intercambiada en un cierto contexto. Así por ejemplo (caballo, yegua, potro, mula) forman un synset porque pueden ser empleados para referirse al mismo concepto. El synset suele ir acompañado de una glosa que describe el significado del mismo.

Otra de las características fundamentales son las relaciones semánticas, en su aspecto de Tesauro, que se establecen entre los synsets tales como hiponimia (entre conceptos específicos y generales), meronimia (entre las partes y el todo), causa-efecto. Es esta estructura de red léxica lo que hace de Wordnet una herramienta versátil para la recuperación de documentos, no sólo textuales, si no en distintos formatos.

Las dos funciones básicas para el empleo de Wordnet en Documentación son la mejora en la Recuperación y su empleo como herramienta de desambiguación en la recuperación de textos e indirectamente para la elaboración de herramientas de clasificación electrónicas, para lo cual se empleará una variedad de estudios de relaciones extraídas a partir de los Corpus de Pruebas.

Wordnet hace uso de una distinción comúnmente aceptada entre las relaciones conceptuales-semánticas, que enlazan conceptos, y las relaciones léxicas que relacionan palabras individuales. La información almacenada en Wordnet para cada término viene dada fundamentalmente por las relaciones léxicas en las que interviene con otros términos. La red (semántica) que es definida por estas relaciones existentes entre las palabras, sirve para dar nombre al sistema.

Esta variedad de relaciones son sistematizadas fundamentalmente en los siguientes apartados:

Relaciones por Materias. Concretamente en Wordnet se agrupan los sustantivos por 23 materias y para verbos se agrupan en 15 materias; aunque para aplicaciones futuras esta agrupación varía de acuerdo con el Corpus al que se aplique.

Relaciones de Sinonimia. La relación de sinonimia se produce cuando dos palabras tienen un significado común; se trata de uno de los principales tipos de relaciones en Wordnet ya que se utiliza como base para la definición del objeto básico en Wordnet: el synset. Por tanto, un synset es un conjunto de sinónimos estrictos. A su vez cada synset en que aparece una palabra representa un significado aparente de la palabra. Los synsets pueden servir como designadores no ambiguos de los dos significados de una misma palabra. De manera intuitiva podemos hacer corresponder un synset con un concepto. Una palabra puede referenciar diferentes conceptos, o synsets. A su vez, un mismo concepto

puede ser referenciado por varias palabras. Por lo tanto el synset sirve como objeto básico en torno al cual se definen las restantes relaciones léxicas en Wordnet.

Relaciones de Jerarquía²⁵⁵. La hiponimia es la relación existente entre conceptos o synsets equivalente a la relación de generalización o del tipo “es-un”. La relación de meronimia da cuenta de la inversa de “tiene-un” o cuando un concepto representado por el synset {x,x',...} es un merónimo del concepto representado por el synset {y,y',...}, en oraciones construidas de la forma “un y tiene un x” o bien “un x es parte de y”. La holonimia se define como relación inversa a la meronimia. A diferencia de la hiponimia y la meronimia, la antonimia se representa a nivel de palabras y no de conceptos.

Relaciones sintácticas. Para cada palabra se incluyen las categorías sintácticas que pueden tener. En concreto el léxico se divide en cuatro categorías: sustantivos, verbos, adjetivos y adverbios, de esta forma los conceptos se encuentran agrupados en cuatro conjuntos disjuntos correspondientes a estas categorías pero la mayor parte de la información almacenada corresponde a los verbos y sustantivos. Lo importante es que para cada categoría se definen distintas relaciones semánticas además de las ya citadas. Por otra parte, la meronimia sólo se define para los sustantivos, y para los adjetivos y adverbios no se define la hiponimia. Para los verbos se definen las relaciones: I) Verbo vinculado por el cual un verbo tiene esta relación con otro si depende su existencia del otro verbo; II) Tropónimo que es un particular tipo de implicación por la cual cada verbo Tropónimo V1 de un verbo más general V2 también implica V2 (por ejemplo: marchar es tropónimo de caminar, pero marchando también implica caminando). III) La relación de Causa que se produce entre un verbo causativo y otro que es llamado resultativo.

Como ya se dijo la función primordial en la recopilación y sistematización de todas estas relaciones en Wordnet es la desambiguación terminológica; pero, ¿en qué campos de la documentación repercutirá el empleo de Wordnet y su faceta de desambiguación? En primer lugar, relacionada con la recuperación de información, permitirá una expansión de la consulta en motores de búsqueda para Internet, para bases de datos y para OPAC's ya que permitirá optimizar la interrogación al sistema y hacer más pertinente la búsqueda y recuperación de información, especialmente de información no textualizada como puede

²⁵⁵ UREÑA LÓPEZ, L. *Resolución de la ambigüedad léxica en tareas de clasificación automática de Documentos*. Alicante: Sociedad Española para el Procesamiento del Lenguaje Natural, 2002.

ser imagen y sonido; ya que la indización en documentos no textuales no se hace a texto completo sino mediante palabras clave, con las que Wordnet puede jugar mediante los distintos tipos de relación para la recuperación exhaustiva y precisa de estos documentos. Otra función, en relación con la clasificación, será la estructuración y categorización de documentos para una nueva organización y representación del conocimiento, especialmente para la búsqueda y recuperación facetada de documentos multimedia mediante la extracción de rasgos semánticos por categorización gramatical de las cuatro categorías léxicas recogidas por Wordnet.

3.3.2.2. El proyecto Framenet.

Algunos lingüistas y lexicógrafos discuten que un análisis puramente relacional no es suficiente para describir la representación que los hablantes se hacen del léxico de los verbos. Por ello, se propuso una teoría basada en marcos cognitivos o esquemas de conocimiento que abarque las experiencias de los hablantes y proporcionen un fundamento conceptual para el significado de las palabras²⁵⁶. Esto significa que aunque las palabras y los conceptos no están directamente interrelacionados, comparten elementos comunes dentro de unos esquemas o marcos fijados de antemano.

Fillmore propone un diccionario conceptual en el cual los significados de las palabras están enlazados con estructuras cognitivas, o conceptos, cuyo conocimiento se presupone para los conceptos expresados por los ítems léxicos. Un ítem léxico está asociado con una descripción de valencia, lo cual permite especificar al mismo tiempo el aspecto sintáctico y semántico de la palabra en su contexto. Fillmore particularmente discute el caso del concepto transacción comercial. Así, distingue las categorías (Dinero, Comprador, Vendedor, Bienes) y los verbos asociados con este concepto y muestra como distintas oraciones y construcciones sintácticas referentes a sucesos de transacciones comerciales pueden ser derivados de una manera sencilla. Aunque Wordnet no enlaza verbos con sustantivos específicos, proporciona marcos sintácticos para cada verbo, para indicación del número de sustantivos argumento que subcategoriza el verbo. Sin embargo, el sustantivo incluido en estos marcos sintácticos no es un objeto enlazado a cualquier función temática o categoría semántica como Comprador. Wordnet podría verse incrementado de tal modo

²⁵⁶ BAKER, Collin F. ; FILLMORE, Charles J. ; LOWE, John B. The Berkeley FrameNet project. En *Proceedings of the COLING-ACL*, Montreal, Canada. 1998.

que las funciones semánticas y temáticas están enlazadas a los sustantivos argumento en los conceptos que acompañan a cada synset. En ese caso la relación de antonimia entre pares de verbos como “comprar” o “vender” podrían verse reflejados en la inversión de los sustantivos argumento también (los argumentos Comprador y Vendedor que son proyectados sobre el sujeto y objeto indirecto en el caso de comprar, respectivamente, son proyectados sobre el objeto indirecto y el sujeto en el caso de vender). Además, comprar y pagar está relacionado en Wordnet mediante enlazando los sustantivos argumento especificados en los marcos sintácticos a funciones temáticas que mostrarían claramente que el argumento Comprador en un caso es también el argumento que expresa la persona que paga.

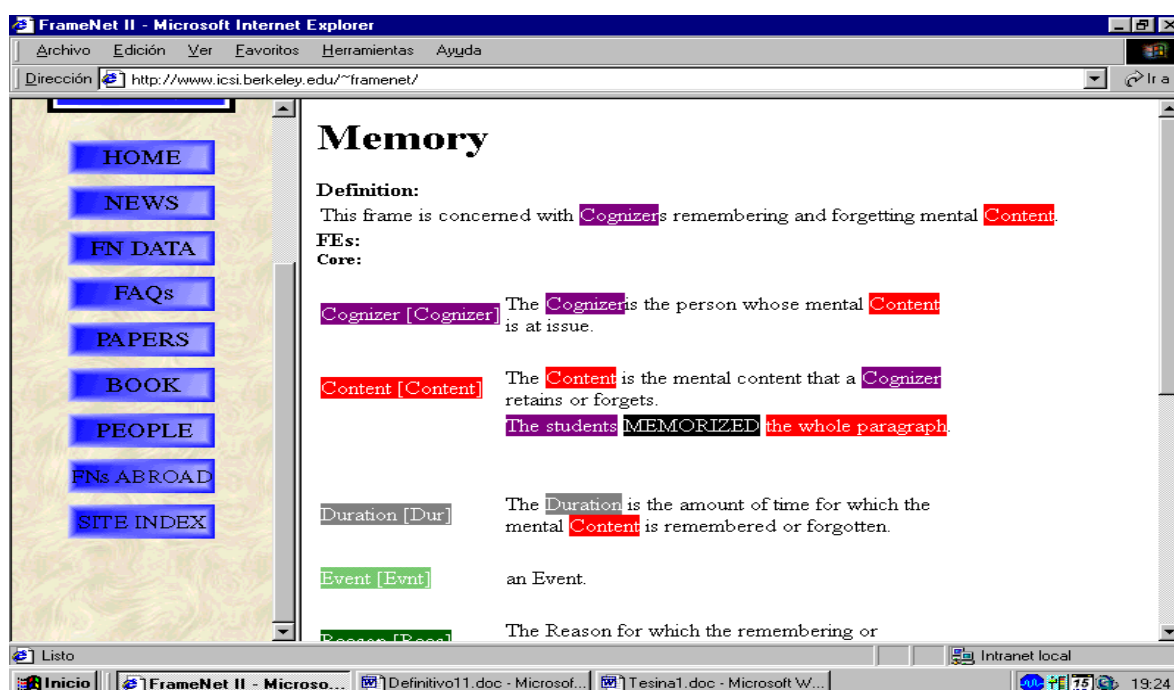


Figura 12. Ejemplo de una búsqueda con Framenet.

El proyecto Framenet²⁵⁷ surgió en la Universidad de Berkeley se hizo accesible al público en 2001. Se trata de una base de datos léxica que destaca por proporcionar, para una parte significativa de términos del inglés contemporáneo, un corpus de sentencias anotadas sintáctica y semánticamente desde las cuales extraer información que pueda ser presentada de acuerdo con las valencias o posibilidades combinatorias de cada elemento que destinado al análisis en la cual se han estudiado palabras del inglés para describir las estructuras conceptuales que subrayan éstas; examina sentencias, empleando un gran corpus de inglés

²⁵⁷ FRAMENET PROJECT. En: <http://www.icsi.berkeley.edu/~framenet/>. Consultado el 15/12/2009.

contemporáneo²⁵⁸ que contiene estas palabras; y registra los modos en el cual la información de los conceptos asociados son expresados en estas sentencias. Éste léxico está orientado a la investigación sobre procesamiento del lenguaje natural orientado a la creación de herramientas para:

- desambiguado del sentido de las palabras: ya que Framenet proporciona información sintáctica y posicional que en muchos casos es probable que esté asociada con un único sentido;
- traductores automáticos: este léxico revelará, mediante el cruce multilingüe de las palabras, las diferencias de significados y comportamientos gramaticales de las palabras que pertenezcan a conceptos análogos.
- extracción de información: Las anotaciones en Framenet proporcionan una alta precisión de búsqueda para la construcción de modelos de extracción de información.
- respuesta a consultas: los datos de Framenet facilitarán el reconocimiento de relaciones conceptuales semánticas entre el lenguaje de interrogación de los usuarios y aquel de los textos de los documentos que responden a la consulta.

El elemento básico de esta base de datos léxica es el frame (concepto o marco). Un concepto²⁵⁹ es un constructo intuitivo que nos permite formalizar los enlaces entre semántica y sintáctica obtenidos como consecuencia del análisis léxico. Los conceptos semánticos son representaciones esquemáticas de situaciones que implican distintos elementos, propósitos y otras funciones conceptuales, cada una de las cuales es un elemento conceptual. Los argumentos semánticos de una palabra del predicado corresponden a los elementos conceptuales del concepto o conceptos asociados con esa palabra.

²⁵⁸ Puede verse un ejemplo práctico de aplicación de Framenet para documentos jurídicos en: Fillmore, Charles J. and Baker, Collin F. Frame Semantics for Text Understanding. En *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL*, Pittsburgh, 2001.

²⁵⁹ PETRUCK, M. Frame Semantics. En VERSCHUEREN, J. ; ÖSTMAN, J. ; BLOMMAERT, J. ; BULCAEN, Ch. (eds.). *Handbook of Pragmatics*. Philadelphia: John Benjamins, 1996.

El trabajo sobre un nuevo concepto comienza con el juicio intuitivo del analista sobre el hablante nativo que algún modelo conceptual particular destaca en una o más unidades léxicas en el lenguaje en un modo sistemático. Nuestros juicios han confirmado que cuando vemos que podemos analizar sentencias conteniendo las palabras que hemos asignado al concepto en términos de los elementos conceptuales que creemos caracterizan ese concepto.

En la vida real, la experiencia actual es un tipo de zigzag cognitivo: un concepto es propuesto, y se sugieren algunas palabras para representar ese concepto. Los elementos conceptuales sugerirían que el analista piensa que estas palabras podrían tener algo en común. Sin embargo, en el curso de la realización de la anotación descubrimos distinciones entre las palabras que en un primer momento no se hicieron evidentes., y a menudo una versión bastante evidente del concepto es formulada.

Decidimos que estamos tratando con un nuevo concepto cuando creemos que las palabras que estamos examinando pueden ser adecuadamente anotadas en términos de las distinciones propuestas para el concepto indeciso, y que no podrían ser tan significativamente anotadas en términos de algún otro concepto en nuestra lista de conceptos. Hasta cierto punto es una cuestión de ensayo y error.

3.4. La “Web Semántica” jurídica.

La mayor parte del contenido disponible actualmente en la web fue diseñado para que cualquier tipo de usuario pudiera leerlo, al mismo tiempo que acceder al mismo a través de las herramientas de búsqueda comunes en la red: motores de búsqueda y directorios. Sin embargo, en muchas ocasiones los usuarios son incapaces de satisfacer sus demandas informativas de manera efectiva a causa de la escasa descripción de los contenidos de las páginas web y de la dificultad para adaptar las consultas del usuario a los contenidos de la web. Especialmente cuando se trata de dominios especializados que pueden ser de interés para el común del ciudadano, como es el caso del derecho.

Sobre este tema aparentemente abstruso Berners-Lee, James Hendler y Ora Lassila publicaron en 2001 un artículo en el que exponían la necesidad de ordenar la información en la web, para lo cual se hace necesario describir sus contenidos. Es ahí donde por primera

vez se define la Web Semántica, como una página cuyos contenidos no sólo son entendibles por las personas sino también por los ordenadores, a fin de que éstos sean capaces de asistir al ciudadano en la adopción de estrategias de búsqueda y selección de los contenidos más adecuados a sus demandas informativas.

La red se ha convertido no sólo en un poderoso medio para la difusión de información y transmisión de conocimiento, sino en un espacio virtual para la interacción entre los ciudadanos y de éstos con las instituciones. A pesar de sus cualidades, su enorme potencial se encuentra limitado por la capacidad del usuario para navegar sin perderse entre los infinitos recursos informativos a los que se proporciona acceso. El aumento exponencial del volumen de información jurídica disponible en Internet, y el bajo porcentaje de precisión y cobertura en su recuperación han originado una necesidad tecnológica en dos sentidos amplios: a) mejorar la web ampliándola con metadatos de carácter semántico que extienden la capacidad de la web hasta lo que se denomina ‘web semántica’; b) crear sistemas de recuperación de información más ágiles y eficaces que sean capaces de entender y gestionar la información de forma ‘inteligente’.

El primer problema que se plantea es el actual lenguaje de marcado de las páginas web jurídicas, HTML, que expresa la forma de presentación de los contenidos del documento y no hace explícita su información semántica. Como resultado los buscadores actuales, tanto de indización manual, Yahoo, como de indización automática, Google, no son capaces de recuperar por el significado de los contenidos transmitidos por los documentos recuperados. Un buscador sólo busca por cadenas de caracteres pero no por el significado de las mismas. Por otra parte, dificulta que el usuario sea capaz de adaptar sus necesidades informativas a los recursos contenidos en la web; es fácil recuperar archivos por su título, pero no por significados abstractos o por materias.

Topic Maps y Tecnologías de la Información Jurídica: un modelo de recuperación de información para bibliotecas digitales semánticas jurídicas.

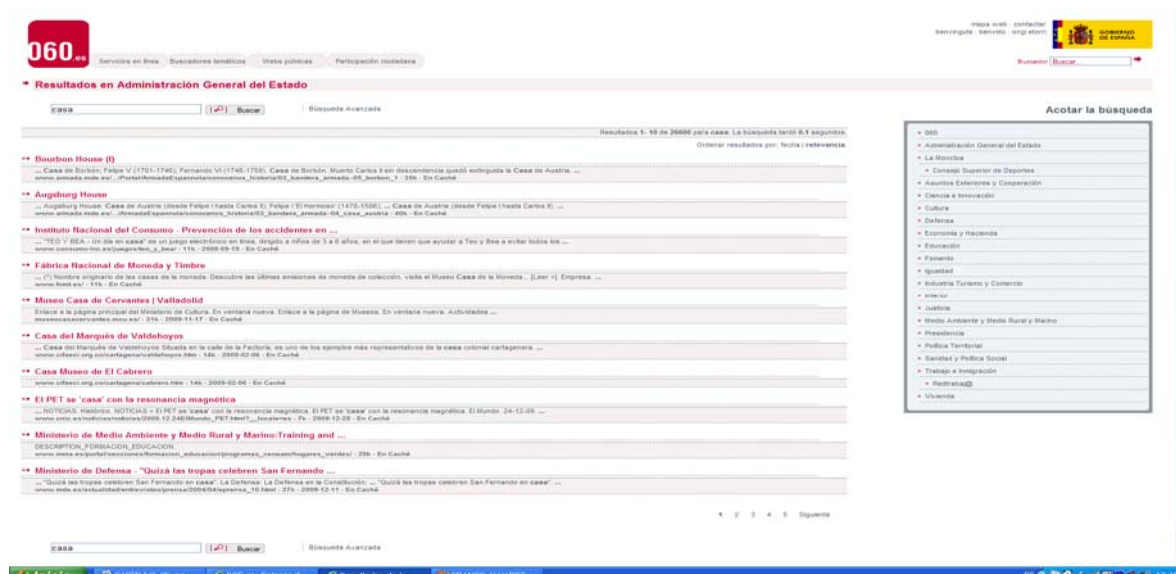


Figura.13. Ejemplo de búsqueda por la palabra “casa” en el portal de la administración 060.es.

Los sistemas de recuperación de información jurídica (SRIJ) pueden representar el contenido semántico de los documentos en esta nueva web a través de las ontologías, que se conciben como recursos idóneos para representar el conocimiento de un área de conocimiento determinado de forma explícita, y son capaces de almacenar la información semántica jurídica mediante los conceptos y las relaciones que se establecen entre ellos. La meta final del desarrollo de ontologías jurídicas es la mejora de la representación de la información y de la efectividad de los SRIJ. La idea que subyace tras la web semántica jurídica es añadirle un nivel de significado, de tal modo que la información que alberga la Red pueda ser interpretada y manipulada más fácilmente por los SRIJ y ‘agentes inteligentes’ de búsqueda.

Por tanto, la web semántica jurídica, es una extensión de la web actual en la que la información recibe un significado bien definido, permitiendo que los ordenadores y los ciudadanos trabajen en cooperación. Se basa en la idea de que los datos de la web se encuentran definidos y relacionados de tal modo que ésta pueda emplearse de manera más efectiva en el hallazgo, la automatización, la integración y la reutilización de dichos datos en diferentes aplicaciones. Para que los sitios web jurídicos puedan alcanzar plenamente su potencial, deben evolucionar hacia la web semántica jurídica, proporcionando una

plataforma accesible universalmente que permita a herramientas automatizadas y a personas compartir y procesar los datos jurídicos²⁶⁰.

En este avance de la web, las ontologías jurídicas surgen como un recurso ideal para que la recuperación de información legal se haga de un modo más preciso. Precisión que se busca a priori mediante la inserción en los sitios web de algún schema de metadatos jurídicos²⁶¹, que hoy en día se busca amoldar a las necesidades de cada institución mediante ontologías y lenguajes controlados de vigencia anterior a la aparición de las ontologías²⁶². Esto permite una descripción contextual del marco legal que puede representarse computacionalmente en un programa²⁶³.

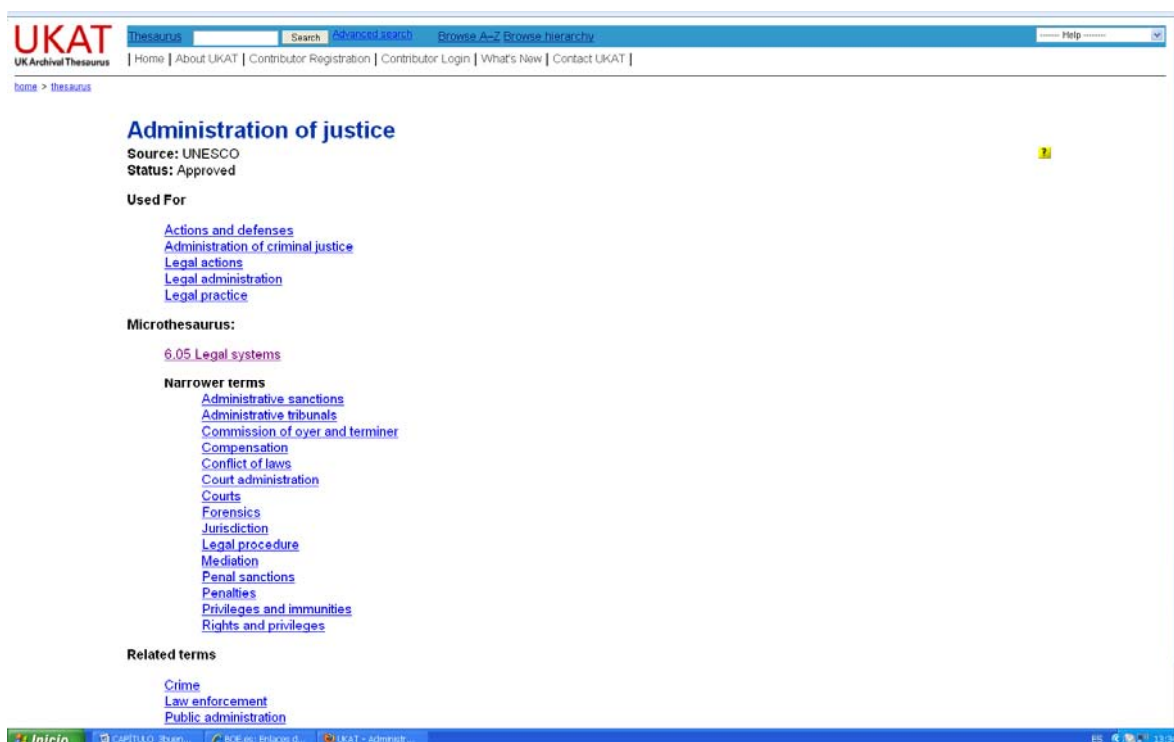


Figura.14. Ejemplo de búsqueda por el tesoro de la administración pública británica: UKAT.

²⁶⁰ HENDLER, J. ; BERNERS-LEE, T. ; MILLER, E. Integrating applications on the Semantic Web. *Journal of the Institute of Electrical Engineers of Japan*, vol.122, 10, 2002, p.676-680. En: www.w3.org/2002/07/swint. Consultado el 12/12/2009.

²⁶¹ Como es el caso del schema de metadatos húngaro para la administración pública KIKERES. En: <http://unpan1.un.org/intradoc/groups/public/documents/CAIMED/UNPAN019384.pdf>. Consultado el 15/12/2009.

²⁶² Es el caso del Tesoro para la administración pública del Reino Unido UKAT. En: <http://www.ukat.org.uk/thesaurus/term.php?i=72> Consultado el 15/12/2009.

²⁶³ STUDER, S. ; BENJAMINS, R. ; FENSEL, D. Knowledge engineering: principles and methods. *Data and Knowledge Engineering*, 25, 1998, pp.161-197. En: <http://citeseer.ist.psu.edu/225099.html>. Consultado el 12/12/2009.

Desde 1990, las ontologías han adquirido un papel destacado en los campos de la ingeniería y la representación del conocimiento. No obstante, y a pesar de que los investigadores procedentes de estas dos disciplinas se revelen como los promotores de esta nueva línea de investigación, es cierto que las ontologías son también relevantes para otras áreas, como las tecnologías del lenguaje humano, el procesamiento del lenguaje natural, las ciencias de la información, la recuperación de información y las bases de datos. Campos todos ellos que de una u otra forma influyen en el desarrollo de la informática jurídica documental y la implantación de las ontologías en los entornos virtuales de información jurídica.

3.4.1. Las tecnologías de la información en la “web semántica” jurídica.

Janes Grey enunció cuáles deberían ser los objetivos de la investigación en tecnología de la información. Fundamentalmente propuso los siguientes objetivos: 1) crear máquinas que se autoreparen; 2) hacer de cada computadora un referencista, es decir una biblioteca ordenada de manera automática e inteligente; 3) producir computadoras inteligentes, esto es que sean capaces de superar el test de Turing²⁶⁴.

Los objetivos de Grey se insertan en la tradición marcada por Vannevar Bush al imaginar el Memex, un dispositivo personal que extendiera la memoria del usuario, capaz de almacenar todos los libros, grabaciones y comunicaciones de una persona. Seguidamente, por medio de un algoritmo, de Turing²⁶⁵ cuando diseñó su máquina, se podría reproducir cualquier conducta humana compleja y creativa; siempre y cuando esta pudiera descomponerse en un número finito y discreto de pasos. En este contexto, se trata de crear sistemas altamente seguros, autoprogramables, autoadministrables y autoreplicables. El dispositivo Memex de Bush se ha materializado, en gran medida, en la web; pero ahora, se requieren sistemas que organicen de forma automática, indiquen y resuman información tal y como lo hacen los documentalistas. Actualmente, las tecnologías de la información tratan de extender el proyecto de las máquinas inteligentes de Turing.

²⁶⁴ El test de Turing consiste básicamente en lograr que un humano no perciba que está interactuando con una máquina sino que suponga que lo está haciendo con otro humano.

²⁶⁵ Consiste en un algoritmo capaz de reproducir cualquier conducta humana compleja y creativa siempre y cuando pueda descomponerse en un número finito y discreto de pasos.

Ya en 2001, Tim Berners-Lee²⁶⁶, advertía que Internet estaba lejos de su madurez y proponía, con el objeto de mejorar la red a nivel global, facilitar la integración de distintos lenguajes web para trabajar conjuntamente, dotarla de funciones móviles de conexión con diferentes dispositivos²⁶⁷. También habló de cuestiones técnicas como la necesaria aprobación oficial de la especificación XML como estándar del consorcio W3C y la aprobación de nuevos estándares como XHTML, SVG o CSS²⁶⁸.

Sin embargo, otras tecnologías que recibieron la atención tanto de este consorcio como del consorcio OASIS²⁶⁹ e instituciones normalizadoras como ISO (Internacional Standard Organization), por su proyección entre las empresas fue la RDF (Resource Description Framework) y XTM (XML para Topic maps).

RDF hace compatibles diversas aplicaciones, lo que permite el intercambio de información de una computadora a otra a través de la web utilizando XML para realizar intercambios de descripciones de cualquier recurso web, incluidas aquellas que no utilizan XML. En tanto que XTM permite, mediante el empleo del metalenguaje XML, el intercambio de un tipo de KOS²⁷⁰ entre colecciones documentales.

El propósito final de ambos lenguajes es facilitar la toma de decisiones por parte de los ordenadores en el momento de ejecutar las búsquedas y para ello se hace necesario que toda la información en Internet se encuentre ordenada en algo que Berners-Lee denominó Web Semántica. La mayor parte del contenido en la web fue diseñado para que los humanos seamos capaces de leerla. Así, por ejemplo, el buscador del CEPC nos proporciona una información relativamente exhaustiva en su dominio, pero si apenas puede identificar titulares o enlaces será incapaz de procesar semánticamente la información. El objetivo es dotar de estructura al contenido de sus páginas web, tejiéndolo en un tipo de entorno que sea procesable por agentes de software.

²⁶⁶ Para más información acerca de Berners-Lee y su trabajo sobre la web. En: <http://www.w3.org/People/Berners-Lee/Overview.html#Talks>. Consultado el 19/12/2009.

²⁶⁷ BERNERS-LEE, T. *Tejiendo la red: el inventor del World Wide Web nos descubre su origen*. Madrid : Siglo Veintiuno, 2000.

²⁶⁸ Véase el sitio web del consorcio W3C para más información sobre la evolución de estos estándares. En: <http://www.w3.org/>. Consultado el 13/12/2009.

²⁶⁹ Consorcio de varias empresas para la aplicación de XML en distintas aplicaciones para empresa. En: <http://www.oasis-open.org/home/index.php>. Consultado el 13/12/2009.

²⁷⁰ Knowledge Organization System= Sistema de Organización del Conocimiento.

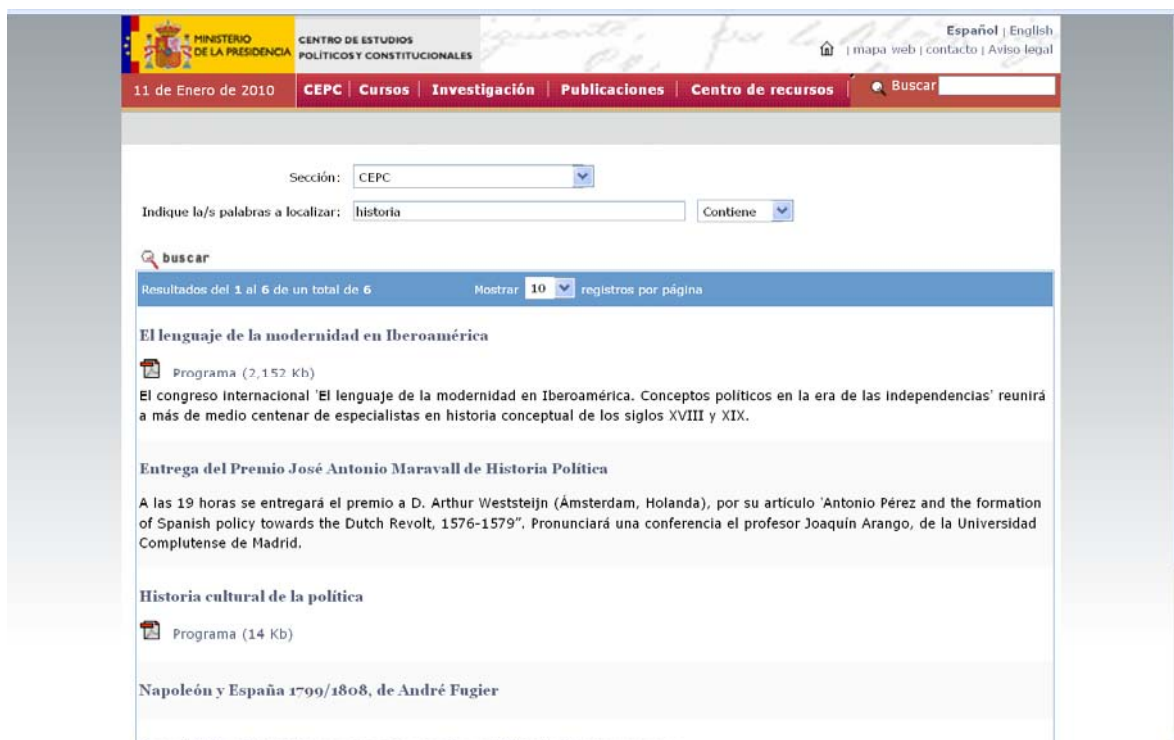


Figura.15. Motor de búsqueda del CEPC para el término Historia.

Por tanto, la evolución de la web consiste en transformar el caos actual propio de la información inconexa y rara vez útil en una estructura de contenidos significativos para la satisfacción de las demandas informativas de los ciudadanos. Se trataría de un entorno especialmente diseñado para que las máquinas se comuniquen entre sí y liberen a los usuarios de tareas documentales para las que normalmente no están preparados y que consumen tiempo²⁷¹.

El rasgo definitorio de la web es su universalidad. Los enlaces hipermediales pueden llevarnos de un medio a otro, pero son incapaces de discriminar tanto por la tipología de los documentos, o la calidad de la información que vinculan como por los posibles usos que pueda dársele. Hecho que se hace especialmente grave en la documentación digital jurídica, donde las tipologías documentales juegan un papel esencial en la correcta comunicación de los actos jurídicos²⁷². Sin embargo, donde la desconexión es aún mayor, entre la información web y el proceso de comunicación web está en la falta de adecuación de los sistemas de información jurídicos para ponderar la pertinencia de los documentos por su semántica, que acompañado de la existencia superficial de páginas HTML inconexas

²⁷¹ BERNERS-LEE, T. Op.cit.

²⁷² MACIÁ, M. *Manual de documentación jurídica*. Madrid : Síntesis, 1998, pp.122-123.

de la información contenida en la web profunda o invisible²⁷³, hace de la “web semántica” jurídica una necesidad perentoria. A esta situación se ha llegado a causa de que la web se había desarrollado como un medio para acceder a documentos y no a datos e información procesada automáticamente; y es aquí donde entran los postulados y herramientas de la inteligencia artificial para comenzar a trabajar de forma automática con todo tipo de datos procesables por máquina y susceptibles de enriquecer los ya existentes en la red.

Un campo con especial incidencia en la web semántica es el referido a la arquitectura de información, es decir la forma en que se organiza, estructura y recupera la información, de tal forma que el usuario pueda acceder a una colección de recursos digitales seleccionados y controlados por expertos en un dominio de conocimiento. Para que la web semántica funcione es necesario que las computadoras tengan acceso a grupos estructurados de informaciones y a conjuntos de reglas de inferencia que les puedan servir para ejecutar razonamientos automatizados, si bien el actual nivel de la arquitectura de información, especialmente en cuanto a la representación de información, se encuentra en un nivel insuficientemente desarrollado como para poder actuar en la web de manera plenamente eficaz²⁷⁴.

Uno de los factores de desarrollo de la web semántica se encuentra en el hecho de que durante mucho tiempo, los programas de inteligencia artificial se orientaron a centralizar las habilidades cognitivas en el cerebro del robot. Como para ello se necesitaba un enorme poder de procesamiento, los sistemas de representación del conocimiento pertenecientes a esas máquinas obligaban a que todo el mundo compartiera los conceptos comunes a fin de no generar problemas de consistencia. Sin embargo, con la actual descentralización de la web se hace factible la creación de la web semántica. Para ello, se parte del supuesto de la ambigüedad como elemento constitutivo de la experiencia cotidiana, en el sentido de poder ofrecer al usuario múltiples posibilidades de navegación por los contenidos de la web²⁷⁵, especialmente si trata de contenidos multimedia.

Así, el poder de la “web semántica” jurídica no está en su estructuración y organización, sino en su expresividad. Para ello, se hace necesario aplicar una lógica que permita el

²⁷³ Constituida por bases de datos, programas e información no recuperada por los motores de búsqueda tradicionales.

²⁷⁴ GEROIMENKO, V ; CHEN, Ch. *Visualizing the Semantic Web*. London: Springer, 2005.

²⁷⁵ MOREIRO GONZÁLEZ, J.A. *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural*. Gijón : Trea, 2004, p.204.

empleo de reglas de inferencia que hagan posible la contestación a preguntas de manera automatizada. Es por ello que su gran desafío es la consecución de un lenguaje que recoja tanto los datos como las reglas de razonamiento de cualquier sistema de representación del conocimiento jurídico que pueda ser exportado a la web²⁷⁶.

La tecnología marco que se ocupará de esta tarea es el metalenguaje XML, dado que permite que cada autor-editor de contenidos web indique sus propias etiquetas, lo cual es equivalente a decir que puede estructurar y organizar el discurso como mejor le convenga. A partir de esta tecnología se han ido desarrollando lenguajes normalizados para la descripción, organización y recuperación de información contextualizada.

Entre estos lenguajes, ya hemos hablado de RDF porque permite construir proposiciones que asignan propiedades a los objetos, donde a cada propiedad se le asigna unos valores, de forma que cualquier recurso de la web, con independencia del tipo que sea, es descrito por medio de este lenguaje. RDF codifica el sentido en un conjunto de tripletes, que son el equivalente computacional del sujeto, el verbo y el predicado de una oración básica. Donde el sujeto y el objeto se identifican por medio de una URI, y el verbo es el elemento de relación. Así, los tripletes constituyen información per se acerca de entidades relacionadas, donde las URI's codifican la información de un documento. Las URI's aseguran que los conceptos no son sólo palabras en un documento, sino que están unidos a una definición única que todos podemos encontrar en la web. Por ello, es tan importante en la “web semántica” jurídica evitar el empleo de metáforas, homónimos y sinónimos en la asignación de URI's.

No obstante, no basta sólo con esto para diferenciar contenidos en Internet. Así como dos bases de datos pueden tener diferentes identificadores para un mismo campo en dos bases de datos, de forma que si las comparamos deberemos contar con algún sistema capaz de informarnos de que ambos modos de designación se refieren a los mismos contenidos. En principio sería deseable que las propias bases de datos fueran capaces de descubrir esos significados comunes cada vez que los localiza en distintas bases de datos. Para solucionar este problema, la “web semántica” jurídica acude a las ontologías jurídicas. En principio,

²⁷⁶ CAMPOS PARDILLOS, M.A. El lenguaje de las ciencias jurídicas: nuevos retos y nuevas visiones. En ALCARAZ VARÓ, E. ; MATEO MARTÍNEZ, J. ; YUS RAMOS, F. *Las lenguas profesionales y académicas*. Barcelona : Ariel, 2007, p.155.

entenderemos la ontología jurídica como un documento o archivo que define formalmente las relaciones entre los términos.

Por medio de las ontologías jurídicas es posible expresar gran cantidad de relaciones entre entidades, asignándoles propiedades a las clases y permitiendo que las subclases hereden esas propiedades. Esto hace posible formular axiomas que podemos definir en un modelo de Chen para el diseño de bases de datos. De este modo, la ontología permite que un ordenador sin entender lo que procesa, pueda manipular los términos de forma eficiente.

Las ontologías jurídicas pueden mejorar el uso de la web en muchos sentidos. Para empezar, pueden al menos facilitar las búsquedas haciendo que los buscadores se fijen en páginas concretas y no en palabras ambiguas. Mucho más interesante es el caso en que las ontologías pueden vincular informaciones, en una página, a las estructuras de conocimiento y a las reglas de inferencia asociadas. Un ejemplo de ello sería al aplicar una ontología al sitio web del CEPC. Al dirigirnos a esta dirección, veremos la página HTML estándar, y como humanos veremos un link a “Centro de recursos” donde sabemos o intuimos que nos llevará a la página web de la biblioteca. Esto que nosotros hacemos espontáneamente sería una tarea compleja para una computadora, que para obtener esa información necesitaría de un programa muy complejo. Por eso los ordenadores, en lugar de ver esa página acceden a una página de ontología que define la información acerca de las distintas secciones y departamentos del CEPC, y al hacerlo realizan las inferencias y vinculaciones que les cuesta obtener a las máquinas. Además, el lenguaje de las ontologías, como OWL, permite desarrollar programas, denominados agentes, capaces de resolver consultas sofisticadas cuya respuesta no está en una sola página. Son programas que negocian entre sí información una vez que sus autores han especificado los términos y alcance de su comportamiento, siempre y cuando se maneje un único lenguaje que asegure esos intercambios.

Por otro lado, el verdadero potencial de la “web semántica” jurídica se verá realmente cuando las instituciones dispongan de programas recolectores de contenido web proveniente de distintos sitios web institucionales, procesen la información y la intercambien, brindando resultados para otros programas. La eficacia de ese software de recolección automática de información crecerá exponencialmente a medida que cada vez haya más contenido legible para las máquinas y servicios automatizados.

Alguna de las posibilidades más importantes que la “web semántica” jurídica tendrá en el desarrollo del derecho son²⁷⁷:

Desde los sitios web jurídicos se va a poder llevar a cabo el intercambio de “pruebas” escritas en un lenguaje unificador²⁷⁸ de la web semántica, OWL, o bien XTM.

Otro rasgo fundamental del sistema serán las firmas digitales; es decir, bloques encriptados de datos que serán utilizados por las computadoras y los agentes para verificar que la información adjunta ha sido brindada por una fuente específica confiable. Los programas agente deben ser muy escépticos acerca de lo que leen en la web semántica, hasta tanto hayan podido chequear exhaustivamente las fuentes de información.

Los agentes del consumidor y del productor pueden llegar a entenderse intercambiando ontologías que brindan el vocabulario necesario para la discusión. Los agentes incluso pueden inventarse nuevas capacidades de razonamiento al descubrir nuevas ontologías.

Se pondrá a disposición de los usuarios la creación de “cadenas de valor”, por los cuales subconjuntos de información pasan de un agente a otro, añadiendo valor informativo a una colección documental hasta constituir el producto final requerido por el usuario.

3.4.2. El lenguaje jurídico en la “web semántica” jurídica.

Deseamos en este epígrafe situar el papel del lenguaje jurídico en relación con la Web Semántica en general y las ontologías en particular. La representación del conocimiento por medio de ontologías es fundamental que se haga con el concurso de la terminología. Cualquier estudio acerca del empleo de lenguajes controlados para la representación del conocimiento cuenta con estudios de terminología, y más si trabajamos con dominios de conocimiento especializado²⁷⁹.

²⁷⁷ BENJAMINS, V.R. ; CASANOVAS, P. ; BREUKER, J. ; GANGEMI, A. *Law and the semantic web, an introduction*. Berlin : Springer-Verlag, 2005, p.5.

²⁷⁸ Lo que hace posible las inferencias lógicas a través del uso de reglas, tal como especifiquen las ontologías.

²⁷⁹ En el ITRI, instituto de investigación en lingüística computacional donde disfruté de una estancia de investigación financiada por el Ministerio de Educación, se trabajaba mucho con Corpora Lingüísticos y el empleo de tesauros y ontologías lingüísticas como Wordnet para la recuperación automática de textos.

La terminología recibe una atención cada vez mayor por parte de lingüistas y especialistas en inteligencia artificial, puesto que es el campo de conocimiento dedicado a estudiar los elementos léxicos que forman parte de un sistema estructurado de conceptos, sistema en donde los términos son los principales responsables de vehicular el conocimiento de una determinada comunidad técnico-profesional.

La terminología es una disciplina transdisciplinar e interdisciplinar²⁸⁰. Transdisciplinar porque la terminología representa un papel primordial en todas las materias especializadas, dado que en ellas se hace uso de la terminología para representar y transferir el conocimiento. Interdisciplinar porque está constituida por elementos procedentes de otras disciplinas, y porque establece estrechas relaciones con áreas del conocimiento como la inteligencia artificial o la Documentación o la lingüística de corpus. Esta naturaleza interdisciplinar viene determinada tanto por las características de las unidades terminológicas, que son al mismo tiempo unidades del lenguaje y vehículos de información, como por las herramientas empleadas en su aplicación práctica en el caso de la informática. A estas disciplinas se deben añadir las nuevas disciplinas de organización del conocimiento y arquitectura de información, así como el desarrollo de aplicaciones para la web semántica como son las ontologías.

Según Temmerman²⁸¹, en la última década se han producido al menos tres cambios relevantes en la disciplina terminológica: uno de naturaleza teórica, otro de corte tecnológico y un último de tipo ontológico. El primer giro se refiere a cuestiones teóricas acerca de la naturaleza de los términos, que empiezan a ser consideradas unidades lingüísticas e informativas, por oposición a la postura generalmente asumida de que los términos no son unidades del lenguaje natural y se caracterizan por ser unidades de conocimiento.

El segundo giro tiene que ver con un nuevo modo de procesar los términos en la actividad práctica de la terminología, pues la gestión terminológica se vincula de forma determinante a las herramientas informáticas destinadas para ello. Gracias al concurso de técnicas, herramientas, y métodos procedentes de campos relacionados con la informática y la

²⁸⁰ CABRÉ, M.T. Op. cit. pp. 22-23.

²⁸¹ TEMMERMAN, R. ; KNOPS, U. Introduction. The translation of domain specific languages and multilingual terminology management. *Lingüística Antverpiensia New Series*, 3, 2004, p.14-15.

lingüística, la gestión terminológica ha experimentado un avance cualitativo en cuatro grandes aspectos. En primer lugar se accede más rápida y fácilmente a documentos e información a través de Internet. En segundo lugar, y gracias a este acceso a la información, se agiliza y facilita la compilación de un corpus de textos digitales representativo del ámbito de especialidad sobre el que se trabaja. En tercer lugar, a partir de este corpus es posible extraer conocimiento especializado utilizando los extractores semiautomáticos de terminología. Y en último lugar, la manipulación, almacenamiento y recuperación de los datos terminológicos se lleva acabo utilizando los datos informáticos de gestión terminológica, también denominados sistemas gestores de bases de datos terminológicas, o sencillamente bases de datos terminológicas. Se trata de herramientas diseñadas específicamente para la gestión de datos terminológicos. Así, la información procedente del material textual se organiza en forma de registros de base de datos, cada uno de los cuales contiene toda la información relativa a cada término distribuida en diversos campos.

En esta última década, cuatro grandes conceptos de naturaleza tecnológica han protagonizado el panorama terminológico: Internet, los corpora digitales, las herramientas para la gestión y extracción terminológica, y las bases de datos terminológicas.

El tercer giro es el ontológico, que combina la representación semiformalizada de las relaciones semánticas con los sistemas informáticos de gestión terminológica, todo ello encaminado hacia la creación de lo que podría denominarse bases de datos termontológicas. En efecto, al considerar las ontologías en el ámbito de la lingüística aplicada, la función más palpable es la de recurso útil para la disciplina terminológica, pues son representaciones formales y explícitas de la estructura conceptual de campos de conocimiento. En consecuencia, dentro de la terminología, y desde la óptica de la terminografía computacional, una ontología se plasma en un módulo vinculado a una base de datos terminológica, creándose así una base de conocimiento, cuya función es proporcionar significado a las unidades léxicas. Es decir, los términos son objetos lingüísticos contenidos en una base de datos terminológica, y dichos objetos se relacionan a través de una red de relaciones conceptuales que está contenida en una ontología, formando una base de datos termontológica.

Tradicionalmente, la investigación terminológica ha prestado atención especial a los conceptos y a sus relaciones dentro de un campo o estructura de nociones. De esta forma,

el concepto no se encuentra aislado sino que forma parte de un área de especialidad y se relaciona con otros, constituyendo así una estructura conceptual o conjunto sistematizado de conceptos que describe una esfera del conocimiento. La estructura conceptual constituye una representación de la realidad del ámbito objeto de estudio. Esta representación pretende recoger y organizar todas las ramificaciones que son propias de un área, de modo que se refleje en forma de esquema la realidad del contexto en cuestión. Dicha representación constituye el sistema de conceptos del área, “el conjunto de conceptos entre los cuales o existen o se han establecido relaciones formando así un todo coherente²⁸²”.

El giro que aquí se produce tiene que ver con el surgimiento de bases de conocimiento terminológicas, capaces de representar y gestionar el conocimiento, además de los términos. Una ontología junto con un conjunto de instancias individuales de clases constituye una base de conocimiento. La creación de este tipo de base de datos supone estudiar los términos tal y como se utilizan en los textos y descubrir las relaciones semánticas que existen entre ellos, también a partir del análisis del material textual. En una base de conocimiento, los conceptos se organizan en redes complejas que representan relaciones como la genérico-específica o de hiponimia, parte-todo o de meronimia, causa-efecto, y objeto-función.

La terminología persigue describir un conjunto de términos de un ámbito especializado. Para conseguirlo, el trabajo terminológico suele seguir un proceso cíclico y constante dividido en fases. Dentro de cada fase se realizan tareas destinadas a sistematizar el trabajo que se define para dicha fase. Las tareas son múltiples y muy variadas; algunas de tipo conceptual y otras de orden documental, textual o lingüístico. En cualquier caso, buena parte de estas aproximaciones necesitan de una aproximación conceptual pues, pues es preciso entender el ámbito que se explora para poder llegar a estructurarlo, clasificarlo y definirlo. Para ello, los especialistas de un dominio de conocimiento ayudan a los terminólogos a estructurar su área de especialidad en forma de sistema de conceptos. Esta delimitación resulta oportuna porque a medida que se clasifica el conocimiento especializado se va haciendo explícita una determinada visión cultural y científica de la realidad. El plano conceptual también está presente cuando se desea entender los términos

²⁸² ARNTZ, R. ; PICT, H. *Introducción a la terminología*. Madrid : Fundación Germán Sánchez Ruipérez, 1995, p.103.

y sus distintas relaciones con otros términos, así como sus significados. En síntesis, una estructura de conceptos debe permitir:

1. Circunscribir la investigación, ya que serán considerados los campos y subcampos en los que el área se divide.
2. Prever tanto el conjunto de términos pertinentes del dominio que se explora, como los que forman parte de ámbitos conexos.
3. Determinar la relevancia de los términos en el dominio, separándolos por grupos referidos a un subcampo dado se puede llegar a establecer cuáles son los verdaderamente interesantes según los fines y el usuario prototípico de la aplicación terminológica que se pretende construir.
4. Elaborar una terminología más controlada y coherente, dado que la estructuración hace posible un tratamiento más sistemático de un área de especialidad.
5. Definir las unidades terminológicas de manera lógica y sistemática.
6. Además es necesaria la recuperación de textos propios de la especialidad para compilar un corpus representativo, y se han de aplicar determinados métodos que son propios de la documentación. Asimismo, al trabajar con textos, será necesario conocer los tipos textuales prototípicos del ámbito y especificar su procesamiento a nivel informático. Los terminólogos están capacitados para localizar, compilar y procesar las fuentes de información necesarias de dominios especializados.
7. Finalmente hay que capturar los términos, definirlos apropiadamente y contextualizarlos; y aquí el terminólogo trabaja a un nivel eminentemente lingüístico, si bien no puede desvincularse totalmente del plano conceptual.

Un creador de ontologías que parte de cero debe de organizar conceptualmente un dominio para descubrir sus conceptos más relevantes y las distintas relaciones que se establece entre ellos²⁸³. Asimismo debe proporcionar definiciones sistemáticas de los

²⁸³ SPYNS, P. ; DE BO, J. Ontologies: a revamped cross-disciplinary buzzword or a truly promising interdisciplinary research topic? *Lingüística Antverpiensia. New Series* 3, 2004, p.285-288.

conceptos, coherentes y desprovistas de ambigüedad, así como codificar dichas definiciones en un lenguaje formal para que las aplicaciones informáticas puedan intercambiar mensajes comprensibles para los ordenadores. Como veremos la creación de ontologías se puede efectuar de manera manual o por procesos automáticos, pero la investigación convergente de la inteligencia artificial, la documentación y la informática es hacia su construcción semiautomática a partir de corpus de especialidad²⁸⁴.

Se observan ciertas similitudes entre la creación de terminologías y la de ontologías. En sus respectivas disciplinas, los dominios explorados se organizan conceptualmente y se acomete la tarea de extraer los términos más relevantes y sus relaciones con otros términos a partir de textos, así como elaborar definiciones coherentes y sistemáticas. Asimismo, en la elaboración de terminologías y ontologías se combinan de forma recursiva dos aproximaciones para capturar y representar el conocimiento de un ámbito especializado: el deductivo y el inductivo. En el primero, en colaboración con los expertos del dominio, se intenta construir la organización conceptual del ámbito mediante la identificación inicial de los conceptos básicos o núdulos primarios que constituyen el punto de partida en la identificación de otros elementos de la ontología. En el segundo se parte de las unidades especializadas insertas en los textos para llegar a sus significados y para ubicarlas dentro de la estructura conceptual del dominio analizado.

Dadas las similitudes, no es extraño que haya surgido un nuevo modo de proceder en la práctica terminológica. Nos referimos al proceso de creación de ontologías como paso previo a la construcción de la base terminológica. Esta nueva metodología ha sido denominada como termontografía²⁸⁵. Al ser un área nueva todavía se está investigando acerca de la especificación detallada de las fases de creación de una ontología de dominio, ya sea manualmente o con la ayuda de extractores de términos. Asimismo se necesita ahondar en el diseño de herramientas informáticas que integren en un único entorno todos los módulos necesarios para la termontografía basada en corpus. Al integrar las ontologías dentro del proceso de trabajo terminológico se hace necesario especificar cómo se elaboran las aplicaciones terminológicas multilingües basadas en ontologías de modo que se plasme

²⁸⁴ DEPRÉS, S. ; SZULMAN, S. Construction d'une ontologie formelle à partir d'un texte de droit communautaire. *Terminology*, vol.11, n° 2, 2005, pp.261-281.

²⁸⁵ TEMMERMAN, R. ; KERREMANS, K. Termontography: Ontology building and the sociocognitive approach to terminology description. Hajicová, E. ; Kotesovcová, A. ; Mirovský, J. En *Proceedings of CIL17*. Praga : Matfyzpress, 2003. En: http://www.hf.uib.no/forskingskole/temmerman_art_prague03.pdf. Consultado el 14/12/2009.

de manera efectiva toda la información contenida en la base de datos termonológica en un esquema de representación del conocimiento útil para la comunidad de usuarios de ese dominio de especialidad.

3.4.3. Ontologías y tipos de ontologías jurídicas.

La filosofía define el concepto de ‘ontología’ como la parte de la filosofía que trata del ser en general y de sus propiedades trascendentales²⁸⁶. La ontología, como campo especializado de la filosofía, se ocupa de la naturaleza y organización de la realidad, e intenta encontrar respuestas para preguntas sobre las características comunes a todos los seres²⁸⁷. Es la ciencia de lo que es, de los tipos de objetos y sus estructuras, propiedades, eventos, procesos y relaciones en cada área de la realidad²⁸⁸.

En la actualidad, el término ontología se utiliza en un gran número de disciplinas, entre las que se encuentran la inteligencia artificial, las tecnologías del lenguaje humano, el procesamiento del lenguaje natural, la representación del conocimiento, la recuperación de información, el diseño de bases de datos, y todas aquellas actividades que tienen que ver con el conocimiento y el lenguaje. Surgen en consecuencia nuevas definiciones para el concepto de ontología, pero la definición más conocida y citada por los autores es la propuesta por Gruber, para quien una ontología es una especificación explícita de una conceptualización²⁸⁹. Otros autores amplían esta definición afirmando que una ontología es una especificación explícita y formal sobre una conceptualización consensuada²⁹⁰. Esta definición supone: 1º) El término ‘conceptualización’ alude a un modelo abstracto de algún fenómeno del mundo procedente de la identificación de los conceptos más relevantes de dicho fenómeno. 2º) Por ‘explícita’ se entiende que los conceptos y demás elementos de una ontología se definen de forma expresa. 3º) El adjetivo ‘formal’ se refiere a que la ontología debe materializarse en algún lenguaje informático computable por el ordenador. 4º) ‘Consensuada’ refleja la noción de que una ontología representa un conocimiento

²⁸⁶ Definición de la RAE. En: www.rae.es. Consultado el 14/12/2009.

²⁸⁷ GUARINO, N. ; GIARETTA, P. Ontologies and knowledge bases: towards a terminological clarification. Toward very large knowledge bases. En *Knowledge building & Knowledge Sharing*. IOS Press, 1995, p.26.

²⁸⁸ SMITH, B. Ontology, 2003. En: http://ontology.buffalo.edu/smith/articles/ontology_pic.pdf. Consultado el 12/12/2009.

²⁸⁹ GRUBER, T. R. A translating approach to portable ontology specifications. *Knowledge Acquisition*, nº 5, 1993, pp. 199-220. En: http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html. Consultado el 12/12/2009.

²⁹⁰ STUDER, S. ; BENJAMINS, R. ; FENSEL, D. Knowledge engineering: principles and methods. *Data and Knowledge Engineering*, nº 25, 1998, pp.161-197. En: <http://citeseer.ist.psu.edu/225099.html>. Consultado el 12/12/2009.

compartido por sistemas diferentes y aceptado por el grupo o la comunidad a la que se refiere.

Esta definición de ontología, si la aplicáramos a la terminología y los lenguajes de especialidad, supone que una ontología pasa a ser una representación formal y explícita de la estructura conceptual del ámbito profesional sobre el que se trabaja. Esta representación del conocimiento de un ámbito dado toma la forma de una red de nodos que representan los conceptos, unidos mediante arcos, que simbolizan las relaciones semánticas que se establecen entre dichos conceptos. Este conjunto de conceptos y las relaciones entre ellos se reflejan en el léxico con el que se representa el conocimiento compartido y común que se tiene sobre algo, pues no hay que olvidar que con el léxico se hace posible la representación de la realidad²⁹¹. El vocabulario de una ontología dada sería el encargado de definir y describir las entidades, clases, propiedades, funciones y relaciones entre los componentes de una ontología.

Las características que habitualmente se atribuyen a los términos se relacionan con el proceso de ‘dar nombre a las cosas’ en la conceptualización²⁹², o proceso de construcción de conceptos, que establecen entre ellos diferentes tipos de relaciones. Este aspecto explica el hecho de que en la historia de la terminología los científicos mostraran un notable interés en establecer nomenclaturas y taxonomías científicas con el objeto de fijar el saber científico y hacer que éste circulase con las máximas garantías de precisión conceptual. En este sentido, la terminología es el reflejo formal de la organización conceptual de una especialidad, y un medio inevitable de expresión y de comunicación profesional²⁹³. El que la organización conceptual de un dominio se haga a través de su terminología refuerza el hecho de que el conocimiento sólo puede ser producido y conocido a través de su materialización en los diferentes sistemas semióticos, entre los que el verbal es el predominante en los ámbitos profesionales y académicos. Por otra parte, se pone de relieve que la confluencia de estas funciones hace aflorar la idea del carácter poliédrico de la unidad terminológica²⁹⁴, constituyéndose al mismo tiempo en un elemento de producción

²⁹¹ ALCARAZ VARÓ, E. ; HUGHES, B. *El español jurídico*. Barcelona : Ariel, 2002, p.17.

²⁹² REY, A. *La terminologie: noms et notions*. Paris : Presses Universitaires de France, 1979.

²⁹³ CABRÉ CASTELLVÍ, M.T. *La terminología. Teoría, metodología, aplicaciones*. Barcelona : Ediciones Antártida, 1993, p. 37.

²⁹⁴ CABRÉ CASTELLVÍ, M.T. *La terminología. Representación y comunicación. Una teoría de base comunicativa y otros artículos*. Barcelona : Institut Universitari de Lingüística Aplicada. Barcelona : Universitat Pompeu Fabra, 1999.

de conocimiento y un recurso de expresión lingüística que favorece la transmisión de dicho conocimiento en contextos de comunicación especializada.

En el ámbito de la inteligencia artificial, cuando el conocimiento de un dominio se expresa mediante un formalismo declarativo, el conjunto de objetos que se pueden representar se denomina ‘universo del discurso’²⁹⁵. Así, el universo de discurso de una ontología es el conjunto de objetos que están representados en ella. Con el fin de representar el conocimiento de algún dominio, las ontologías se componen de los siguientes elementos²⁹⁶:

Clases, conceptos o topic types. Son el punto neurálgico de las ontologías, pues se trata de las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias o procesos de razonamiento.

Subclases o topics. Constituyen en sí mismas una clase pero con la característica de ser dependientes de alguna otra clase.

Relaciones o associations. Representan la interacción y el vínculo que se establece entre los conceptos del dominio. Suelen formar la taxonomía del dominio.

Funciones o roles. Son un tipo concreto de relación por la que se identifica un elemento con otra función que comprende varios elementos de la ontología.

Instancias u occurrences. Se utilizan para representar objetos determinados de una subclase o concepto.

Axiomas. Son teoremas que se formulan sobre las relaciones que deben cumplir los elementos de la ontología. En topic maps, estas normas se contienen en la sintaxis TMCL (Topic map Constraint Language).

Desde una perspectiva terminológica, el universo de discurso está constituido por los textos especializados que se producen y circulan en un dominio dado y que conforman un

²⁹⁵ GRUBER, T. R. A translating approach to portable ontology specifications. *Knowledge Acquisition*, nº 5, 1993, pp. 199-220. En: http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html. Consultado el 12/12/2009.

²⁹⁶ LOZANO TELLO, A. Ontologías en la web semántica. *Cuadernos de investigación en ingeniería informática*, nº5, 2001. En: www.informandote.com/jornadasIngWEB/articulos/jiw02.pdf. Consultado el 14/12/2009.

corpus de textos a partir del cual se procede a la extracción terminológica. En una ontología²⁹⁷, las definiciones vinculan los nombres de las entidades en el universo de discurso con un texto legible por una persona que describe lo que los nombres pretenden denotar, y con axiomas formales que restringen la interpretación y el uso gramaticalmente correcto de dichos términos. Una ontología está compuesta por una taxonomía de términos, y por definiciones y axiomas que especifican y restringen la interpretación de los términos en un dominio dado.

En cuanto a los tipos de ontologías que es posible construir, están en función de parámetros tales como el ámbito de conocimiento al que se apliquen, según el tipo de agente al que vayan destinadas, y según el grado o nivel de abstracción y razonamiento lógico que permitan.

En función del ámbito de la conceptualización se pueden clasificar en²⁹⁸:

- a) Ontologías genéricas o de alto nivel. Proporcionan términos generales reutilizables en diferentes dominios. Así, términos como espacio, tiempo, estado, evento, acción los ofrecen este tipo de ontologías y sirven para todo tipo de ontologías.
- b) Ontologías de dominio o de tareas. Expresan conceptualizaciones que son específicas de dominios concretos. Los conceptos en las ontologías de dominio se definen habitualmente como especializaciones de conceptos existentes en las ontologías de alto nivel. Resultan útiles para dominios concretos con un lenguaje y ámbito muy específicos como puede ser el campo del derecho.
- c) Ontologías de aplicación. Contienen todas las definiciones que son necesarias para dar forma a los conocimientos necesarios para una aplicación particular. Incluyen conceptos tomados de las ontologías de dominio y de las genéricas, a menudo definidas utilizando el vocabulario indicado en las ontologías de representación.

El derecho constituye un ámbito privilegiado de aplicación de las web ontologías, ya que en el ámbito jurídico se necesita contar con estructuras conceptuales compartibles, escalables y

²⁹⁷ GRUBER, T. R. A translating approach to portable ontology specifications. *Knowledge Acquisition*, nº 5, 1993, pp. 199-220. En: http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html. Consultado el 12/12/2009.

²⁹⁸ GÓMEZ-PÉREZ, A. ; FERNÁNDEZ-LÓPEZ, M. ; CORCHO, O. *Ontological engineering*. London : Springer, 2004, p.29-30.

reutilizables²⁹⁹ y éstas son características de las ontologías. Existen ya algunas ontologías jurídicas³⁰⁰ que han permitido la construcción de prototipos de asistencia, búsqueda y organización documental de la información almacenada en bancos de datos y portales jurídicos³⁰¹.

El vínculo entre lo que el jurista tienen mente y los documentos jurídicos se denomina ontología jurídica. Por tanto, la ontología jurídica es la representación de la estructura de los objetos conceptuales del jurista que le permite compartirlos con la red de bases de datos jurídicas. Se trata de representaciones del conocimiento terminológico que constituyen la base semántica para los servicios inteligentes de la Web Semántica.

Las ontologías jurídicas comenzaron a desarrollarse bastante antes de la aparición del concepto de Web Semántica, con la idea de ser empleadas para la gestión del conocimiento jurídico y como bases de conocimiento en los sistemas jurídicos³⁰². Sin embargo, no existen tecnologías específicas del ámbito jurídico que articulen una Web Semántica jurídica. Más bien, el desarrollo de la misma depende de la actividad de la comunidad de personas e instituciones que demandan un conjunto de servicios de valor añadido que podemos situar en el ámbito de la Web Semántica. Además, en la edición de sistemas de información jurídica, la estructuración del ámbito regulatorio es una tarea esencial para muchas áreas de aplicación de la Web Semántica: e-comercio, e-gobierno, e-política.

Las ontologías jurídicas difieren de las ontologías de otros campos en que deben de cubrir un abanico más amplio de conceptos que son parte de otros dominios de conocimiento como la ciencia política, la sociología, etc. Asimismo, los distintos dominios jurídicos comparten nociones variadas y complejas de normas con matices divergentes según el momento y el ámbito de origen y aplicación de las normas. No obstante, existe un núcleo básico de nociones de teoría del Derecho que las ontologías jurídicas deben incluir, tales

²⁹⁹ CASANOVAS, P. Ontologías jurídicas profesionales. Sobre “conocer” y “representar” el derecho. En: http://www.leibnizsociedad.org/secciones/mater/pon/textos/ontologias_pompeu.pdf. Consultado el 15/12/2009.

³⁰⁰ Los primeros proyectos europeos se sitúan en el 5º Programa Marco (1999-2003). Comprenden e-POWER, CLIME, e-COURT. Posteriormente se le dieron continuidad a estos proyectos en el 6º Programa Marco, dentro del grupo SEKT (Semantically enabled knowledge technologies)

³⁰¹ CONTRERAS, J. (et al.). A semantic portal for the international affairs sector. En *Proceedings of EKAW 2004: engineering knowledge in the age of the semantic web*. Berlin : Springer, 2004, pp.203-215.

³⁰² BENJAMINS, V. R. Law and the Semantic Web, an Introduction. En *Law and the semantic web: legal ontologies, methodologies, legal information retrieval, and applications*. Berlin : Springer, 2005, p.9.

como: norma, institución, documento legal, responsabilidad, etc.; que para la construcción y mantenimiento de ontologías jurídicas, se ven estructuradas a través de un conjunto de técnicas desarrolladas por la Documentación y la Ingeniería de Ontologías: análisis conceptual, representación del conocimiento.

Existe un número creciente de trabajos de investigación³⁰³ acerca de cómo se construyen y aplican ontologías en el ámbito legal, de lo cual ha surgido en los últimos tiempos un nuevo campo de estudio denominado Ontologías jurídicas³⁰⁴. En la imagen siguiente se muestre el modelo de arquitectura de información IURISERVICE, que emplea una ontología como modelo lógico de procesamiento de las consultas³⁰⁵.

Figura.16. Esquema de una arquitectura de información con ontologías. Fuente ISOCO.

³⁰³ LegOnt 2003. En *Proceedings of the ICAIL 2003 Workshop on Legal Ontologies & Web based legal information management*. Edinburgh, 2003. En: www.lri.jur.uva.nl/~winkels/legontICAILE2003.html Consultado el 12/12/2009.

³⁰⁴ LegOnt 2001. En *Proceedings of the Second International Workshop on legal ontologies*. Amsterdam, 2001. En: www.lri.jur.uva.nl/jurix2001/legont2001.htm. Consultado el 12/12/2009.

³⁰⁵ BENJAMINS, V.R. (et al.). Iuriservice: Un FAQ inteligente para los jueces en su primer destino. En: <http://www.sepln.org/revistaSEPLN/revista/31/31-Pag347.pdf>. Consultado el 15/12/2009.

Las estructuras jurídicas existentes tienen una fuerte relación con los conceptos de la teoría clásica del derecho. Esto significa que entre cualquier ontología de alto nivel y la ontología de dominio se sitúa un nivel interpretativo intermedio que no puede ser obviado. Es en este nivel donde se toman las decisiones teóricas básicas sobre los conceptos cuyas relaciones la ontología va a contemplar para efectuar el vínculo con las instancias del nivel inferior y las categorías fundamentales que asumen las ontologías de nivel superior³⁰⁶.

La representación del conocimiento incorporado a una ontología plantea el reto de reconstruirlo de un modo que permita la identificación del vocabulario común, la organización de la estructura de relaciones entre sus conceptos básicos y la esquematización de las pautas de razonamiento más comunes en el campo del derecho³⁰⁷. La forma de realizar estas tareas en las ontologías jurídicas consiste en la representación formal de los conceptos nucleares, que se expresan en el lenguaje jurídico. Sin embargo, como Casanovas nos dice, existen tres estrategias fundamentales para su representación: a) la construcción o empleo de un lenguaje controlado en el que se expresan relaciones prefijadas entre conceptos previamente definidos para aplicarlo luego a la descripción de los documentos; b) formalizar conceptos fundamentales ya provistos por la tradición jurídica a partir de la combinación de operadores lógicos con axiomas; c) refinar las interrelaciones de una gran cantidad de instancias de un dominio acotado a partir de una generalización conceptual que permita deslindar las diferencias semánticas entre el lenguaje jurídico y el lenguaje natural.

La “web semántica” jurídica se ha ido desarrollando en los últimos años a medida que la iniciativa de la Web Semántica se mostraba como un nivel más de desarrollo de la web tanto por su integración de nuevas técnicas de las tecnologías de la información como por la arquitectura de información, las técnicas de representación del conocimiento o la ingeniería del conocimiento. Sin pretender entrar de lleno en la variada tipología de ontologías jurídicas creadas hasta el momento, vamos a mostrar en una tabla-resumen algunas de las más importantes iniciativas en este campo³⁰⁸.

³⁰⁶ CASANOVAS, P. Ontologías jurídicas profesionales. En:

http://www.leibnizsociedad.org/secciones/mater/pon/textos/ontologias_pompeu.pdf. Consultado el 12/12/2009.

³⁰⁷ Id.

³⁰⁸ VALENTE, A. Types and Roles of Legal Ontologies. En *Law and the Semantic Web*. Berlin : Springer, 2005. pp.65-76

Proyecto de Ontología	Aplicación	Tipo	Rol
Ontología funcional de derecho de Valente y Breuker³⁰⁹	Ofrece una arquitectura general para la resolución de problemas legales.	Repositorio de conocimiento en Ontolingua altamente estructurado.	Comprensión del dominio, razonamiento y resolución de problemas jurídicos.
Modelo Jurídico basado en conocimiento de Mommers³¹⁰	Lenguaje general para la expresión de conocimiento legal	Base de conocimiento en inglés poco estructurada	Comprensión de un dominio.
Ontología marco de Van Kralingen y Visser³¹¹	Lenguaje general para la expresión de conocimiento legal y sistemas basado en conocimiento legal.	Representación de conocimiento moderadamente estructurado, incluso como base de conocimiento en Ontolingua.	Comprensión de un dominio.
Ontología de McCarty para discursos legales³¹²	Lenguaje general para la expresión de conocimiento legal	Representación de conocimiento altamente estructurada	Comprensión de un dominio.
Ontología para motivaciones legales de	Generalmente para la expresión de conocimiento legal	Base de conocimiento en inglés poco	Comprensión de un dominio.

³⁰⁹ VALENTE, A. ; BREUKER, J. Legal Modelling and Automated Reasoning with ON-LINE. *International Journal of Human-Computer Studies*, vol.51, n°6, 1999. pp.1079-1125.

³¹⁰ MOMMERS, L. *Applied legal epistemology. Building a knowledge-based ontology of the legal domain*. Leiden, 2003. En: https://openaccess.leidenuniv.nl/dspace/bitstream/1887/4432/1/Proefschrift_Mommers.pdf. Consultado el 25/12/2009.

³¹¹ VISSER, P.R.S. *Knowledge specification for multiple legal tasks. A case study of the interaction problem in the legal domain*. The Hague : Kluwer Law International, 1995.

³¹² MCCARTY, L.T. A language for legal discourse, I. Basic Features. En *Proceedings of the second international conference on artificial intelligence and law*, pp. 180-189. Vancouver, 1989.

Lehman, Breuker y Brower³¹³	causal	estructurada	
Ontologías de conocimiento legal profesional de Benjamins³¹⁴	Sistema FAQ de recuperación de información de información para juicios	Base de conocimiento en Protégé, moderadamente estructurada	Indización semántica y búsqueda
Ontologías de Lame para códigos franceses³¹⁵	Recuperación de información legal	Base de conocimiento orientada al procesamiento del lenguaje natural, poco estructurada.	Indización semántica y búsqueda
Ontología contra el fraude financiero de Leary, Vandenberghe y Zeleznikow³¹⁶	Ontología para la representación de casos de fraude financiero	Base de conocimiento en UML, poco estructurado	Indización semántica y búsqueda
Ontología de textos legales de Saias y Quaresma³¹⁷	Creación semiautomática de ontologías a partir de texto.	Base de conocimiento léxica en OWL, moderadamente estructurada	Indización semántica y búsqueda

³¹³ LEHMAN, J.; BREUKER, J. ; BROWER, B. Causation in AI&Law. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies & Web based legal information management*, 2003. En: <http://www.lri.jur.uva.nl/~winkels/LegOnt2003/Lehmann.pdf> Consultado el 12/12/2009.

³¹⁴ BENJAMINS, V. R. ; CONTRERAS, J. ; CASANOVAS, P. ; AYUSO, M ; BECUE, M ; LEMUS, L. ; URIOS, C. Ontologies of professional legal knowledge as the basis for intelligent IT support for judges. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies & Web based legal information management*, 2003. <http://www.lri.jur.uva.nl/~winkels/LegOnt2003/Benjamins.pdf> Consultado el 12/12/2009.

³¹⁵ LAME, G. Using text analysis techniques to identify legal ontologies components. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies & Web based legal information management*, 2003. <http://www.lri.jur.uva.nl/~winkels/LegOnt2003/Lame.pdf> Consultado el 12/12/2009.

³¹⁶ LEARY, R. M. ; VANDENBERGHE, W. ; ZELEZNIKOW, J. Towards a financial fraud ontology. A legal modeling approach. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies & Web based legal information management*, 2003. En: <http://www.lri.jur.uva.nl/~winkels/LegOnt2003/Leary.pdf> Consultado el 12/12/2009.

³¹⁷ SAIAS, J. ; QUARESMA, P. Using NLP techniques to create legal ontologies in a logic programming based web information retrieval system. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies & Web based legal information management*, 2003. En: <http://www.lri.jur.uva.nl/~winkels/LegOnt2003/Saias.pdf> Consultado el 12/12/2009.

Jur-Wordnet de Gangemi, Sagre y Tiscornia³¹⁸	Extensión de Wordnet al dominio legal.	Base de conocimiento léxica en DOLCE (DAML), ligeramente estructurada	Organiza y estructura información
Ontología criminal italiana de Asaro³¹⁹	Schema para la representación de crímenes en derecho italiano.	Base de conocimiento en UML ligeramente estructurado	Organiza y estructura información
Ontología CLIME de Boer, Hoekstra y Winkels³²⁰	Sistema de asesoramiento legal para derecho marítimo	Base de conocimiento en Protégé y RDF moderadamente estructurado	Razonamiento y resolución de problemas.
Desarrollador de argumentos de Zeleznikow y Stranieri³²¹	Sistema basado en la combinación de múltiples sistemas legales	Representación de conocimiento moderadamente estructurado	Razonamiento y resolución de problemas.

Tabla 4. Principales proyectos de ontologías jurídicas desarrolladas hasta la fecha.

³¹⁸ GANGEMI, A. ; SAGRI, M. T. ; TISCORNIA, D. Metadata for content description in legal information. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies & Web based legal information management*, 2003. En: <http://www.lri.jur.uva.nl/~winkels/LegOnt2003/Gangemi.pdf> Consultado el 12/12/2009.

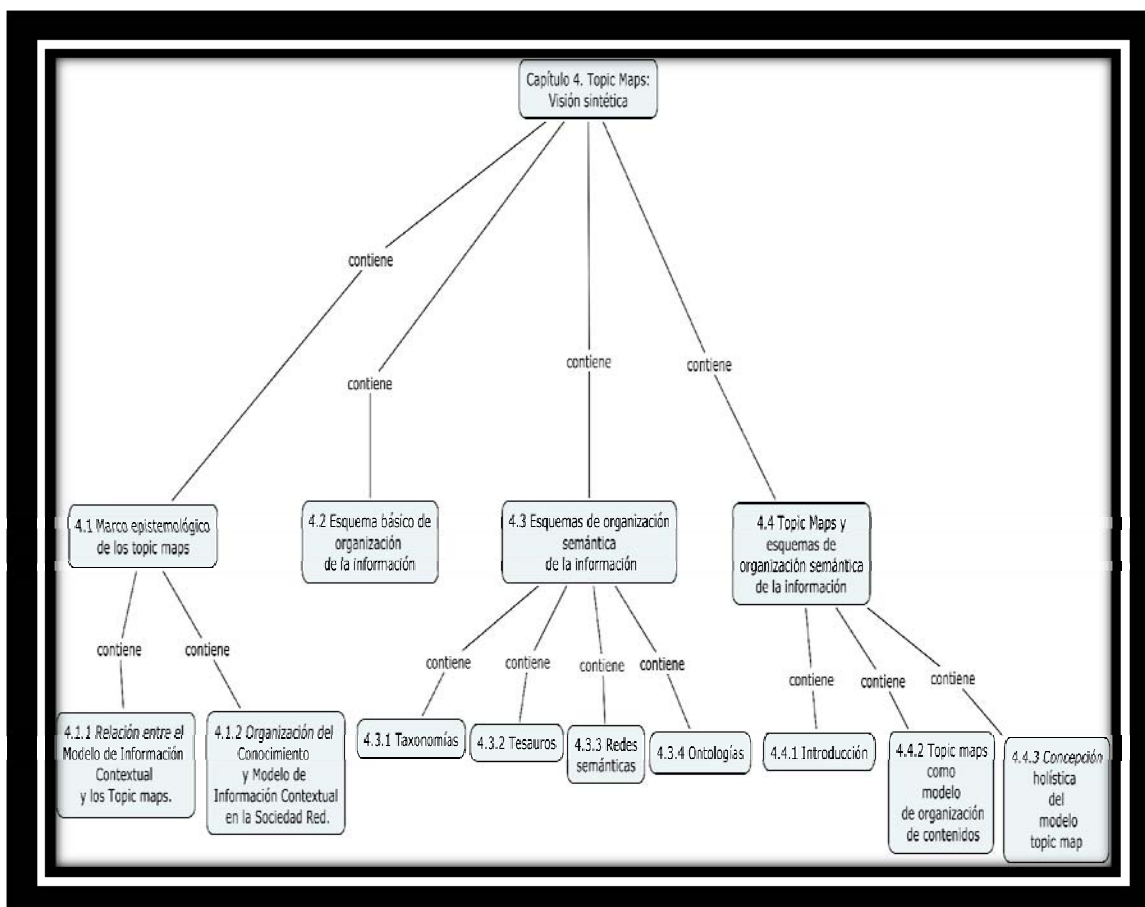
³¹⁹ ASARO, C. ; BIASIOTTI, M. A. ; GUIDOTTI, P. ; PAPINI, M. ; SAGRI, M. T. ; TISCORNIA, D. A domain ontology : Italian crime ontology. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies & Web based legal information management*, 2003. En: <http://www.lri.jur.uva.nl/~winkels/LegOnt2003/Asaro.pdf> Consultado el 12/12/2009.

³²⁰ BOER, A. ; HOEKSTRA, R. ; WINKELS, R. The CLIME ontology. En *Proceedings of the second international workshop on legal ontologies*. Amsterdam, 2001. En: <http://www.lri.jur.uva.nl/jurix2001/papers/Boer.pdf> Consultado el 12/12/2009.

³²¹ ZELEZNIKOW, J. ; STRANIERI, A. An ontology for the construction of legal decision support systems. En *Proceedings of the second international workshop on legal ontologies*. Amsterdam, 2001. En: <http://www.lri.jur.uva.nl/jurix2001/papers/Zeleznikov.pdf> Consultado el 12/12/2009.

CAPÍTULO 4.

Topic maps: Visión Sintética



4.1 Marco Epistemológico de los *Topic maps*.

Nos encontramos en una etapa de la Historia en que hemos pasado de la dificultad para buscar información a la sobrecarga de información, o más bien, como algunos autores matizan, una etapa de angustia informativa³²² en la que se hace imprescindible repensar cómo presentamos la información. Y la solución a esta sobrecarga de información pensamos que está en su estructuración, especialmente si, como nos dice Brown³²³, la información no es más que datos estructurados con un propósito. Pero no es suficiente con organizar datos, ya que sin el concurso de una metodología que nos indique los pasos a

³²² WURMAN, S. *Angustia informativa*. Amsterdam : Prentice-Hall, 2001, p.8ss.

³²³ BROWN, P. *Information Architecture with XML. A management strategy*. Chichester : Wiley, 2003, p.36

seguir en la organización y diseño de un modelo de información no es posible construir conocimiento³²⁴.

Toda estructura de información se basa en una metodología para el diseño de las arquitecturas de información. De forma especial si tratamos con información digital tanto por su carácter hipertextual como por la obsolescencia de sus contenidos o la carencia de un espacio físico singular en donde se pueda hallar el contenido íntegro de una colección³²⁵.

Por otra parte, se hacen necesarios procedimientos de edición adaptados a las recientes necesidades de información de alto contenido semántico. Las metodologías para el diseño de información deben ser desarrolladas y fundamentadas en distintos métodos de navegación y organización, así como en las nuevas herramientas y lenguajes de programación³²⁶.

La sobrecarga de información es una anomalía cotidiana para cualquier usuario de Internet, causada por la estructura multidimensional de la información. En 1945, Vannevar Bush ya pronosticó la posibilidad de llegar a un entorno informativo mal estructurado y de escasa calidad³²⁷ dado el ritmo acelerado de generación de información que se estaba produciendo. Y en efecto, esto es lo que ha sucedido, especialmente desde la aparición de Internet. La edición de información se ha ido acelerando de forma exponencial en los últimos años, lo cual ha llevado a muy distintas ramas del conocimiento a buscar instrumentos de selección y localización de la información, pero todavía no hemos sido capaces de encontrar una solución adecuada al problema³²⁸. De hecho resulta paradigmático el elevado número de motores de búsqueda que se están creando hoy en día para la recuperación exhaustiva y rápida de información por Internet³²⁹; sin embargo, como nos decía V. Bush, ¿qué ha sido de la calidad de la información, de la precisión en las búsquedas, de la pertinencia de la información recuperada en el contexto semántico para el que se recupera?

³²⁴ FISHER, M. Developing an information model for information- and knowledge- based organizations. En GILCHRIST, A. ; MAHON, B. *Information Architecture. Designing information environments for purpose*. London : Facet, 2005, p.6.

³²⁵ Ibid., p.7.

³²⁶ Ibid., p.9.

³²⁷ BUSH, V. As we may think. *Atlantic Monthly*, 1945.

En: <http://www.theatlantic.com/doc/194507/bush>. Consultado el 12/12/2009.

³²⁸ LYMAN, P. *How much information*. En: www.sims.berkeley.edu/research/projects/howmuch-info-2003/ Consultado el 12/12/2009.

³²⁹ De hecho acaba de aparecer un nuevo motor de búsqueda, *CUIL*, que pretende ser más exhaustivo y rápido que Google. En: www.cuil.com. Consultado el 28/12/2009.

En este sentido abogamos por un modelo para la estructuración de información más adecuado a las características de la información digital semántica. Pensamos, como explicaremos a continuación, que el modelo de estructuración de la información de Nordeng y Karabeg, denominado Polyscopic³³⁰, es el idóneo para explicar la integración de la información en un entorno de Web Semántica.

El modelo Polyscopic³³¹ procura resolver el problema de estructuración de la información como paso previo para la solución de las dificultades de localización de información adecuados al contexto semántico de búsqueda.

Hasta ahora la información web ha estado, en general, poco estructurada. El hecho de acceder por medio de un navegador a una página web, se debe a que esta se encuentra organizada a partir de un esquema de etiquetas en HTML³³² que agrupan, de forma más o menos aleatoria, contenidos. Algunos de ellos varían a fin de atraer la atención del usuario, pero en general no hay una organización global de información; es decir, la disposición informativa de la web es caótica, y pocas veces con algún tipo de distribución global. En general, los sitios web son unidades deslocalizadas, sin ningún tipo de estructura que permita conectarlos y organizarlos entre sí³³³. Estos sitios web se conectan entre sí por medio de hipervínculos, similares a las primitivas opciones de “ir a”, por los cuales se pincha en un hipervínculo y somos transferidos a otro sitio web donde se localizan otros hipervínculos con capacidad para transferirnos a sucesivos sitios web. Al final, después de múltiples accesos el usuario es incapaz de comprender cómo ha llegado a donde está y por qué se encuentra allí. A este fenómeno se le conoce por sobrecarga cognitiva³³⁴ y se caracteriza por la necesidad, por parte del usuario, de efectuar un esfuerzo adicional de concentración para poder simultanear varias tareas al mismo tiempo³³⁵.

³³⁰ Se trata de un modelo desarrollado en la Universidad de Oslo por los profesores Dahl y Karabeg.

³³¹ Polyscopic es como lo han designado los autores, pero por el significado que le confieren lo denominaremos contextual, ya que se refieren a una palabra compuesta de: “poly” (=muchos) y scopic o scope (=ámbito, dimensión), lo que proporciona un entorno multidimensional o contextual.

³³² No olvidemos que la Definición del Tipo de Documento o DTD para HTML ya está definida cuando se aplica a la edición de una página web.

³³³ Especialmente si hablamos de web social como blogs.

³³⁴ NORDENG, T ; GUESCINI, R ; KARABEG, D. Topic Maps for polyscopic structuring of information. *International Journal of Continuing Engineering Education and Lifelong Learning*, vol.16, nº1/2, 2006, pp.37-38.

³³⁵ CONCLIN, J. Hypertext: an introduction and survey. *IEEE Computer*, vol. 20, nº9, septiembre 1987, pp. 17-41.

El problema de la sobrecarga cognitiva en Internet es que se ha producido muy poca información de alto nivel³³⁶. Esta carencia de estructuras normalizadas, no sólo capaces de ligar información en el ámbito de un mismo dominio de conocimiento sino de decirnos tanto el significado intrínseco de la misma como si ésta es pertinente o no para satisfacer la demanda informativa puntual de cada usuario, lastra las posibilidades de creación de conocimiento a partir de la información en Internet.

Bush ya nos indicaba que la presentación secuencial de información es poco adecuada a la forma asociativa en que realmente opera la mente humana para la localización de información³³⁷:

La mente humana no funciona de esta manera. Opera por asociación. Con una idea que el hombre comprenda es capaz en el momento de pensar en otras ideas por asociación, de acuerdo con una intrincada web (red) de pistas dirigidas por las células del cerebro. Tiene otras características como el hecho de que aquellas pistas que no son frecuentemente empleadas son propensas a ser progresivamente borradas dado que la memoria es transitoria. Incluso la velocidad para actuar va en función de la complejidad de las pistas. El detalle de las imágenes mentales va más allá de su naturaleza intrínseca.

Nuestra mente y sentidos no sólo nos proporcionan una idea intuitiva de lo que vemos a nuestro alrededor en todo momento, sino que también nos permite seleccionar las ideas más relevantes de acuerdo con un propósito dado. Así uno de los principales desafíos para el modelo de información contextual es trasladar el conocimiento contextual del hombre de la vida real al entorno virtual. Esto significa que debemos ser capaces de organizar la información de tal modo que el usuario tenga una idea clara de la dimensión³³⁸ o contexto en el que se mueve, con una visión clara del sentido de aquello que se conceptualiza.

³³⁶ Entendemos por información de alto nivel como aquella que es necesaria para ligar toda la información que se encuentra dispersa por la red, como puedan ser los metadatos, los schemas o los lenguajes controlados.

³³⁷ BUSH, V. As we may think. *Atlantic Monthly*, 1945. En: <http://www.theatlantic.com/doc/194507/bush> . Consultado el 12/07/2009. Consultado el 12/07/2009.

³³⁸ Entendiendo la dimensión o ámbito de conocimiento por lo que tanto Karabeg y los teóricos de los topic maps designan por scope.

El modelo contextual de la información de Dahl y Karabeg que proponemos para la aplicación de topic maps, establece dos dimensiones para la información, lo que coincide con el modelo bidimensional de Maniez³³⁹. Ambas dimensiones son:

I) La dimensión paradigmática o vertical de la información.

Uno de los problemas más graves de la sobrecarga de información es el relativo a su elevado grado de fragmentación. A medida que se incrementa la información, las personas tienen que especializarse más. Esto les lleva a crear información dentro de sus propios subdominios de especialidad, lo que conduce a la creación de islas virtuales de información. Carecemos de un metanivel de información capaz de conectar todas estas lagunas informativas, capaz de proporcionarnos una visión más general y precisa de los distintos grupos de información de especialidad. En otras palabras, la raíz del problema de la sobrecarga de información está en su escasez a un nivel descriptivo y contextual superior. A un nivel informativo elevado es posible dotarse de una idea intuitiva, o al menos más holística de la información presente dentro de un dominio de conocimiento. Con esto se procura comprender globalmente acerca de qué trata un campo de conocimiento. El modelo de información multidimensional puede ser representado por medio de un ideograma como el siguiente³⁴⁰:

³³⁹ MANIEZ, J. *Actualité des langages documentaires : les fondements théoriques de la recherche d'information*. Paris : Association des professionnels de l'information et de la documentation (ADBS), 2002.

³⁴⁰ Ideograma de la información Polyscopic de Karabeg. GUESCINI, R. ; KARABEG, D. ; NORDENG, T. A Case for Polyscopic Structuring of Information. En Maicher, L. ; Park, J. *TMRA* 2005, pp.125-138.



Figura 17. Ideograma del modelo Polyscopic.

En el ideograma se representa un triángulo dentro del cual está inscrita toda la información, representada por la letra “i” como un cuadrado que sostiene el círculo del punto. El hecho de que ésta sea representada por “i” y se encuentre inscrita en un triángulo significa que existe una jerarquía de dominios o contextos. De esta manera se indica que la información multidimensional está jerárquicamente estructurada. En ella se distinguen fundamentalmente dos partes representadas por el punto de la “i” y por el trazo de la misma. El punto representa la información de alto nivel, y el trazo representa la información de bajo nivel. El usuario emplea la de alto nivel como un tipo de resumen o punto de vista informativo, mientras que aquella de bajo nivel se emplea para detallar la información y al mismo tiempo es la base sobre la cual generar la de alto nivel como metadatos, resúmenes o descriptores de un lenguaje controlado.

II) La dimensión sintagmática u horizontal.

Uno de los principios del modelo polyscopic es que el contexto, el punto de vista desde el que se observa en la pirámide informativa, condiciona la perspectiva del usuario³⁴¹. Este principio exige que al usuario se le proporcione la posibilidad de escoger entre distintas perspectivas que se le ofrecen a partir de la selección del ámbito más adecuado. Esto permite dotarle de un control sobre la manera en que accede a la información.

³⁴¹ NORDENG, T. ; GUESCINI, R. ; KARABEG, D. Topic maps for polyscopic structuring of information. *International Journal of Continuing Engineering Education and Lifelong Learning*, vol. 16, nº1/2, 2006, p.39.

La metáfora que mejor explica la dimensión horizontal en la estructuración de información la tenemos cuando, en la realización de un dibujo técnico, se hace necesaria la representación de la perspectiva. Para ello, es necesario escoger un plano de proyección que muestre el lado de los objetos que deseamos sea representado desde un cierto ángulo, pero esta imagen no representa la totalidad del objeto. Para representar todo el objeto hacen falta varios planos de proyección. Por tanto, en la recuperación hará falta la selección de unos planos de proyección adecuados, *scopes*, para que: a) cada uno de ellos por separado pueda proporcionar un punto de vista de todo el objeto de información, b) todos los planos de proyección juntos sean capaces de representar la totalidad del objeto de información³⁴².

Si fusionamos la dimensión paradigmática y sintagmática estaremos hablando de información holística. De la letra “i” que representa la información inscrita en el triángulo: el punto y el trazo. Luego, de acuerdo con el modelo de información contextual, ésta no estará completa si únicamente es representada por una de estas dos dimensiones, ya que ambas son necesarias.

La información global se encuentra representada tanto por la información de alto nivel como por la de bajo nivel, combinadas, pero distinguibles en ambos niveles. Análogamente al ideograma, así como el rectángulo del trazo sirve de pedestal para el círculo, la información de bajo nivel sirve como fundamento para aquella de alto nivel. Por otra parte, para que toda ella se encuentre completa no es suficiente con proporcionar un conjunto de metadatos o de descriptores, sino que es necesario proporcionar todas las dimensiones o dominios relevantes para cada contexto informativo.

Es así como en el mundo real el hombre no sufre sobrecarga informativa, ya que el aprendizaje le permite seleccionar en cada momento los contextos más adecuados al mensaje que se recibe. En el mundo virtual los ordenadores no están ni adaptados ni programados para la selección de un contexto o *scope* en un momento dado, lo cual genera sobrecarga informativa, dado que presenta toda la información que tenga algún tipo de

³⁴² Ibid., pp.39-40.

relación con la demanda informativa; es decir, hemos sido incapaces de transferir y aplicar el principio natural del contexto al mundo virtual³⁴³.

Para muchos autores³⁴⁴, la organización jerárquica de la información es la manera de afrontar la ausencia de un contexto que obliga a nuestros usuarios a tratar sincrónicamente un todo informativo. Pero si tratamos con pequeñas partes en distintos momentos, entonces el conjunto será mucho más manejable. La organización contextual modular y asíncrona es la metodología que el modelo de información Polyscopic nos ofrece para tratar con la complejidad de las grandes colecciones informativas.

Así como se emplean módulos para simplificar los programas de ordenador, también deben ser empleados para la organización de la información. El módulo de más alto nivel nos proporciona una imagen de conjunto. Los módulos de bajo nivel nos proporcionan los detalles, normalmente ordenados jerárquicamente y posiblemente siguiendo una correspondencia entre los módulos de alto y bajo nivel. Lo esencial de estos módulos es que cada uno de ellos sea coherente, lo que significa que no es posible mezclar una visión global con sus detalles. La imagen global de la colección informativa se presenta de manera diferente en el módulo de alto nivel, mientras que los detalles se presentan en unos módulos de bajo nivel representantes individualmente de un punto de vista coherente, lo cual facilita que la complejidad global de la información se haga manejable.

La razón por la que abogamos por una metodología para el diseño de información digital se debe a que no basta con reunirla y organizarla bajo los mismos principios jerárquicos que se siguen en la organización de la información textual, caracterizada por su secuencialidad. Para crearla, se necesitan metodologías que aún son escasas.

Una metodología de diseño de información proporciona un principio de organización con el cual estructurar una colección de conceptos de acuerdo con los siguientes principios básicos:

i) el diseño de información estructurada se basa en la idea de una jerarquía de módulos que representan los distintos niveles de abstracción y,

³⁴³ Ibid., p.41.

³⁴⁴ WURMAN, R. S. *Angustia informativa*. Buenos Aires : Pearson Education, 2001, p.123ss.

ii) la orientación de los objetos de información se basa en la metáfora del objeto informativo, y la idea de programación en lo referido al modelado de un objeto de la vida real basándose en las funciones del objeto³⁴⁵.

Así, por ejemplo, la metodología contextual proporciona un marco para la organización de información sobre la base de la noción de ámbito. Esta organización, en función del contexto, se entiende fácilmente a partir de la metáfora del embudo, la cual puede tener muy distintas caras o contextos pero que, en cualquier caso, reúne todos los elementos para proporcionar un único resultado final.

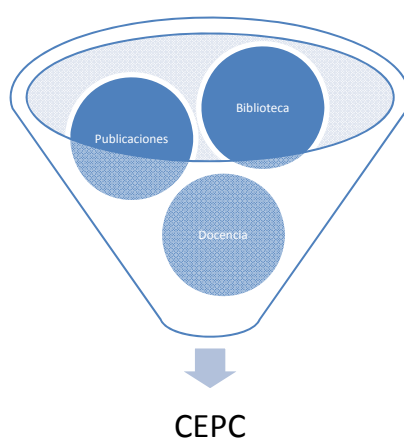


Figura 18. Esquema representativo de la metodología contextual.

En el modelado contextual se proporcionan tres tipos de abstracción: vertical, horizontal, y estructural. Estos tres niveles se corresponden con las cotas de una montaña. Al igual que las vistas desde lo alto de una montaña, la información de alto nivel proporciona una fotografía que nos permite la comprensión de un fenómeno, o la elección del modo con el que afrontar un problema informativo. Mientras que las panorámicas de bajo nivel proporcionan los detalles. Los aspectos corresponden a los ángulos de observación, reflejando las necesidades de los usuarios, las categorías de usuario, o los distintos aspectos de la materia cubierta³⁴⁶.

³⁴⁵ GUESCINI, R. ; KARABEG, D. ; NORDENG, T. A case for Polyscopic structuring of information. En Maicher, L. ; Park, J. *TMRA 2005*, LNAI 3873, p.127ss.

³⁴⁶ *Ibid.*, p.128.



Figura 19. Los tres tipos de abstracción en el diseño de información contextual.

La navegación contextual permite la concepción de la estructura global de un documento en módulos que reflejan distintos niveles y puntos de vista con los que el usuario puede visualizar la información. Este tipo de navegación se basa en el principio de que el scope determina la vista³⁴⁷, lo que en este contexto significa que al usuario se le proporciona la posibilidad de seleccionar una vista sin más que escoger entre los distintos scopes ofrecidos. Intuitivamente, el contexto viene a ser un dispositivo de control para poder efectuar la navegación entre vistas. Únicamente hace falta una adecuada representación visual de la estructura modular, con una indicación clara del scope actual, que le sirva de orientación al usuario.

La metodología del modelo contextual orienta la creación y uso de información proporcionando unos criterios explícitos. Según Karabeg³⁴⁸:

La perspectiva de criterios evidencia que la información necesita permitirnos ver el documento descrito como un todo, en las adecuadas proporciones, con nada esencial oculto.

Una perspectiva clara de la información contenida en un documento significa que cada parte del todo informativo debe ser visible. Esto no significa necesariamente que todos los

³⁴⁷ NORDENG, T. Op. cit., p.42.

³⁴⁸ KARABEG, D. *Information design-a new paradigm in creation and use of information:IPSI*. En: <http://folk.uio.no/poly/ID-paradigm.pdf>. Consultado el 15/12/2009.

detalles sean proporcionados en una sola vista, sino al contrario. Si se proporcionan demasiados detalles, la perspectiva deja de estar clara. El contexto o dimensión adecuada, y particularmente el nivel correcto en la jerarquía contextual, debe ser escogida de forma que la materia sea transparente. El principio que aquí se afirma es que la complejidad de los contenidos está en la mente del usuario, lo cual implica que toda materia puede hacerse evidente sin más que seleccionar el scope adecuado. La información tiene la oculta pero esencial capacidad de alimentar las ideas del hombre, que es lo que le permite construir su conocimiento. En este sentido el criterio de relevancia debe ser priorizado de acuerdo con los propósitos que la información necesite alcanzar.

Este principio aboga por el hecho de que hay ciertas funciones que un sistema de organización del conocimiento puede realizar por el usuario, y una de ellas es dar prioridad a la información relevante y útil. El documentalista debe estar seguro de que la información que es relevante y útil sea la prioritaria para nuestros usuarios, especialmente en un entorno de información jurídico donde se procura que la información sea fiable y verificable³⁴⁹.

Cada módulo de información debe fundamentarse tan sólida y ampliamente como sea posible. Que esté bien fundamentado significa que permite su acceso a un amplio abanico de usuarios con variado rango de juicio tecno-científico. El fundamento es sólido si no puede ser falsable por ningún tipo de usuario, experto o no en la materia³⁵⁰.

Se están desarrollando métodos para la creación de contenidos que satisfagan estos criterios científicos. Para ello, y al mismo tiempo, resolver el problema de la sobrecarga de información se necesita que ésta se encuentre estructurada de forma contextual.

4.1.1 Relación entre el Modelo de Información Contextual y los *Topic maps*.

El modelo de información *topic map* es apropiado para la aplicación de estructuras contextuales de información³⁵¹. Cualquier declaración de un *topic map*³⁵² se efectúa en el

³⁴⁹ KUHN, T. La estructura de las revoluciones científicas. México: Fondo de Cultura Económica, 2001. p.23ss.

³⁵⁰ Ibid., p.27.

³⁵¹ PASSIN, T. Navigating information with topic maps. En *Explorer's guide to the semantic web*. USA : Manning, 2004, p.61.

³⁵² Tal como la asignación de un nombre, una association o una occurrence, cualquiera de estos componentes de los *topic maps* debe ser declarado por medio de una statement que liga cada uno de estos componentes con el resto del *topic map*.

contexto de un scope o ámbito de aplicación. Esto permite aclarar la semántica de la sentencia que liga dos *topics* de forma que pueda ser empleada para la recogida de distintos puntos de vista en el mismo *topic map*³⁵³.

Una de las fortalezas de los *topic maps* es su potencial para crear capas de información de alto nivel en la organización de información dispersa en grandes redes informativas como Internet o grandes bases de datos. En este sentido, el modelo polyscopic se aplica al diseño de información en Internet mediante la aplicación de topic maps. En general, podemos distinguir tres modos de aplicación en función de cómo se desee organizar la información. Estos modos son³⁵⁴:

a) En caso de necesitar crear estructuras para la aplicación de la dimensión vertical, deberíamos tener en cuenta la posibilidad de diseñar módulos capaces de presentar vistas de alto y bajo nivel. Lo primero que se debería generar son las clases o types: “alto” y “bajo” nivel. Así cualquier topic debe ser una instancia de un topic type o una clase topic. Empleando associations como “superclase-subclase”, “clase-instancia”, o algún otro tipo de association adecuado como “parte-todo”, o incluso “alto nivel/bajo nivel”, uno puede fácilmente crear jerarquías y diseñar la estructura principal sobre la cual se basan los niveles taxonómicos. Sin embargo, hay un problema para la creación de presentaciones donde un usuario puede estar interesado en crear *topic types* que gradualmente contengan conocimiento gradualmente más detallado dentro de un dominio; esto podría lograrse por medio del uso de algún tipo de *association type* como puede ser: “más/menos detallado”.

No obstante debemos recordar que aunque creemos topics de bajo nivel, éstos no dejan de ser una abstracción de aquellos de alto nivel para fragmentos de información que están convirtiéndose en occurrences dispersas. Una solución a este problema sería por medio de la creación de topic types de alto nivel que representen los fragmentos de información u occurrences y luego añadirles los fragmentos como occurrences a ese topic o bien sólo rectificar una occurrence como un topic de bajo nivel y adjuntarle metadatos.

³⁵³ KARABEG, D. ; GUESCINI, R. ; NORDENG, T.W. Flexible and exploratory learning by polyscopic Topic-maps. En *ICALT* 2005. En: www.win.tue.nl/SW-EL/2005/swel05-icalt05/final/W3-2.pdf. Consultado el 12/12/2009.

³⁵⁴ NORDENG, T. ; GUESCINI, R. ; KARABEG, D. Op. cit., pp. 38-40.

b) La navegación contextual se basa en el principio de que “el ámbito determina el campo de visión”, y los ámbitos del *topic map* aplican la abstracción horizontal; es decir, los aspectos del modelo contextual³⁵⁵. Una manera de aplicar la abstracción horizontal es por medio de la creación de *topics* de un ámbito definido para cada aspecto necesario en la presentación, y a continuación adjuntarlos a las *associations* empleadas para aplicar la abstracción vertical. Así, este filtrado proporciona al usuario la posibilidad de seleccionar una vista dependiente del ordenador sobre la base de estos *topics* en un ámbito definido.

c) La abstracción estructural trata de la estructura y navegación de los documentos en su contexto, y por tanto se ejecuta a nivel de aplicación. Dado que se debe diferenciar entre módulos de alto y bajo nivel, se debe cuidar el uso de los *topic types*. Cada módulo diferente tiene que ser clasificado de forma que el *topic map* pueda ser procesable por el ordenador a nivel de aplicación. Ya que si éste ya es difícil para un desarrollador de aplicaciones *topic map*, aún lo es más si deseamos ejecutar un modelo automático de información contextual con *topic maps* aplicado a la parte de navegación, donde se produce el problema inherente en los enlaces hipertextuales de falta de estructuración, haciéndose necesario crear buenas estructuras navegacionales para combatir este problema³⁵⁶. Esto se afronta por medio de las *associations*, los *scopes* y los *topic types*. Y como veremos es la combinación del modelo de datos *topic map* y la estructuración contextual como afrontaremos el objetivo de producir una presentación de información holística.

4.1.2 Organización del Conocimiento y Modelo de Información Contextual en la Sociedad Red.

Según Pink³⁵⁷, no hay duda, que hemos pasado de la era de la información a la era de lo conceptual, en una progresión temporal exponencial que es función del incremento de la riqueza, el desarrollo de la tecnología y su nivel de globalización. Esta sucesión de etapas de desarrollo intelectual humano ha ido conformando una sociedad del conocimiento, que Castells ha denominado la sociedad red³⁵⁸.

³⁵⁵ GUESCINI, R. Op. cit. p.130.

³⁵⁶ MARCHIONINI, G. Búsqueda exploratoria: de encontrar a comprender. *Novática: Revista de la Asociación de Técnicos de Informática*, Nº185, 2006, p.7.

³⁵⁷ PINK, D. H. *A whole new mind. Moving from the information age to the conceptual age*. New York : Riverhead books, 2005, p. 49

³⁵⁸ CASTELLS, M. *La sociedad red*. Madrid : Alianza, 2005. pp.34ss.

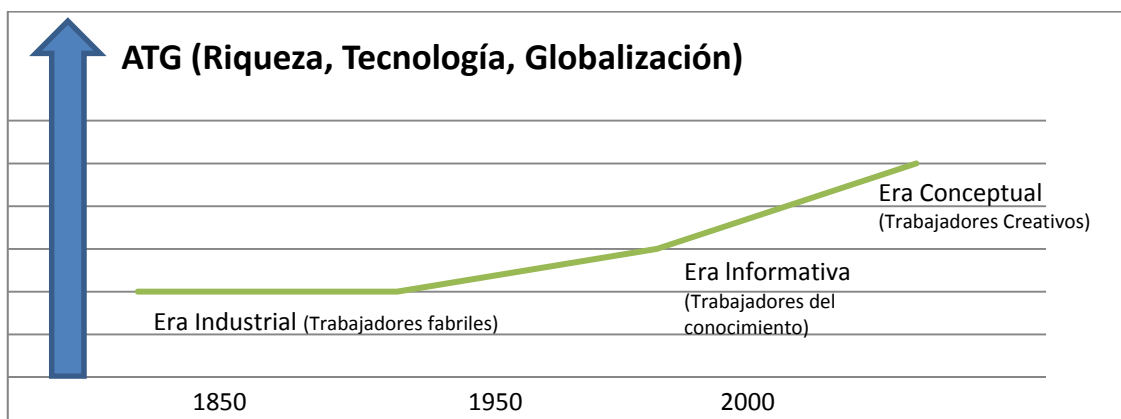


Figura 20. Gráfico de Pink. Representación temporal de las etapas de desarrollo en función de ATG.

La sociedad red que surge a partir de la cotidianeidad de Internet, amplifica y potencia la mente humana, por eso el conocimiento de sus usos y efectos es esencial para adaptar la más poderosa tecnología de información y comunicación a nuestros valores, intereses y proyectos³⁵⁹. Millones de personas navegan diariamente por Internet y buena parte del sistema económico de los países industrializados depende de la información digital almacenada y gestionada por redes de ordenadores. La información digital se duplica cada dos años, haciendo que exista una sobreabundancia de información que dificulta el uso y acceso al conocimiento. Por ello la sociedad del conocimiento se está orientando hacia la difusión de información semántica en la cual no sólo las personas sino también los ordenadores sean capaces de recuperar por significados y no por cadenas de caracteres³⁶⁰.

El peligro, para su consecución, reside en que nosotros no sólo accedemos a un único recurso de información correspondiente a una materia dada, sino que podemos acceder a miles de recursos. La información sobre una materia puede residir tanto en los registros de una base de datos, como en los documentos de una empresa, en partes concretas de algunos documentos, en páginas web, imágenes o vídeos que se encuentran en distintos repositorios o fuentes de información. Además, de manera progresiva, se está haciendo necesario conocer el contexto en el cual se encuentra un recurso de información, a fin de poder ponderar su pertinencia en relación con una consulta; así como aquellos enlaces capaces de permitirnos recuperar los recursos informativos asociados con lo que estamos consultando. Se trata de capacitar al usuario para extraer conocimiento de la información

³⁵⁹ Ibid., p. 46.

³⁶⁰ BERNERS-LEE, T ; HENDLER, J. ; LASSILA, O. The Semantic Web. *Scientific American*, vol.284, nº5, p.35

contextualizada. Esto nos lleva a un objetivo más refinado para la sociedad de la información y es el de su difusión pertinente en el contexto pertinente para la persona pertinente en el momento pertinente, porque la sociedad del conocimiento nos exige pertinencia. En este sentido, la función de la organización del conocimiento es asistir al usuario en la toma de decisiones sobre la base del contexto y la experiencia sobre su entorno³⁶¹.

La selección, no sólo de la información, sino de su contexto más adecuado para las necesidades informativas surgidas de la actividad del usuario constituye una de las principales tareas de toda biblioteca³⁶²; y por supuesto también, de una institución pública consultiva y docente como es el CEPC. Pero es que además, como tal centro gestor del conocimiento, tiene el cometido de acortar todo lo posible el tiempo de aprendizaje organizativo a fin de incrementar la capacidad de adaptación de la organización a las nuevas variables del entorno. Se constata así la existencia de un ciclo de conocimiento basado en: reconocimiento, captura, distribución, utilización y retroalimentación de las materias. Esto permite un aprendizaje en la organización que la faculta como una entidad con una elevada capacidad para explotar el ciclo de conocimiento.

Cuando concebimos la organización del conocimiento desde el punto de vista de la tecnología, se distinguen: i) la gestión de información; y por otro, ii) el acceso a la información³⁶³.

La gestión de la información actúa para asegurar la actualización, precisión y coherencia de la información. Por tanto, su finalidad es proporcionar acceso a una cantidad creciente de información interrelacionada, aunque diversa y en ocasiones muy compleja, capaz de resolver el problema de la sobreabundancia de información.

Para ello, divide los distintos tipos de información gestionada en: a) gestión de datos, b) gestión de activos como imágenes y gráficos, c) gestión de documentos, d) gestión de contenidos tales como documentos estructurados con acceso a las distintas partes del

³⁶¹TAYLOR, A. G. *The organization of information*. London : Libraries unlimited, 2004, p.131.

³⁶² *Ibid.*, pp.231-232

³⁶³ *Ibid.*, pp.32ss.

mismo, e) gestión de metadatos, f) gestión de enlaces entre objetos de información, y g) gestión de estructuras de conocimiento.

El acceso a la información asegura que ésta se encuentre disponible para la persona correcta, a un correcto nivel de detalle de una manera coherente y uniforme; pero, si nos fijamos, el acceso a la información puede también subdividirse en: a) búsqueda (en todos los tipos de información), b) navegación (a través de estructuras de conocimiento), y c) notificación (cuando la información es añadida al sistema de gestión de acuerdo con un perfil de usuario)³⁶⁴.

A partir de la publicación del artículo de Berners-Lee, se formuló la idea de la Web Semántica. El término Web Semántica se refiere a la visión de una web donde no sólo el hombre sino también las máquinas son capaces de extraer inferencias e intercambiar conocimiento a un nivel más alto de lo que actualmente es posible. El tipo de inferencia que se considera es la referida a aquella que permitiría inferir estructuras cognitivas tales como: si el gobierno aprueba un Real Decreto que desarrolla una ley, y esa ley formula un marco jurídico para un ámbito social concreto, entonces el Real Decreto vendría a desarrollar, o modificar la aplicación de esa misma ley. Esto está relacionado con el problema del acceso a la información puesto que este mecanismo mejora la precisión en la recuperación si los documentos jurídicos están anotados semánticamente³⁶⁵ de forma que permita al sistema aplicar las reglas de inferencia.

La solución que plantea la Web Semántica es asignar objetos a la Web que describan la forma y el contenido de la Web y a su vez las relaciones con otras páginas Web. Estos objetos son los metadatos. A fin de asegurar la compatibilidad entre metadatos se hace necesario acordar un modo común de realizar anotaciones semánticas. Es aquí donde los lenguajes controlados realizan un papel clave.

Enriquecer las páginas web con metadatos es una solución al problema de la organización y la recuperación, donde se entiende por metadato en su sentido más amplio de datos acerca de datos. Tener metadatos acerca de un objeto ayuda a encontrarlo puesto que la colección de metadatos para ese objeto funcionará como un sustituto, en el cual los elementos

³⁶⁴ MARCHIONINI, G. *Information seeking in electronic environments*. Cambridge : Cambridge University Press, 1997, pp.112ss.

³⁶⁵ Anotados semánticamente significa que tienen insertas una serie de etiquetas informativas sobre la estructura del texto.

implícitos pueden ser explicitados y por tanto incrementar las posibilidades de lograr una recuperación pertinente y precisa. En este sentido, las distintas aproximaciones a la clasificación y la indización se basan en la necesidad de emplear los metadatos. Así por ejemplo cualquier número de clasificación de la CDU es una parte de los metadatos asignados a una referencia bibliográfica. La noción de metadato puede ser comparada a la idea del catálogo de una biblioteca que contiene archivos donde el aspecto físico de los libros, la información editorial tal como autor y clasificación por materia es registrada. Otra ventaja del empleo de metadatos es que un sustitutivo es más fácil de tratar que el objeto en sí mismo y puede ser manipulado de forma más flexible que con el objeto físico.

El concepto de metadato es independiente del medio en el cual está registrado, pero en el caso de los medios electrónicos se deben tener en cuenta los siguientes aspectos para crear y usar de forma efectiva los metadatos.

Qué metadatos deben ser registrados. Esto implica que se necesitan definiciones estructurales y de contenido acerca de los metadatos. Así por ejemplo podríamos considerar si usamos o no MARC. MARC es el formato de catalogación standard, aunque con distintas variantes todas ellas acordes con la norma ISO. Una iniciativa mas reciente es Dublin Core, desarrollado desde mediados de los 90 y desde 2003 norma ISO. Dublin Core consiste de una colección de 15 campos de metadatos lo suficientemente generales como para poder describir cualquier tipo de objeto digital. La verdadera fuerza del esquema de metadatos Dublin Core reside en su generalidad puesto que sus 15 campos pueden ser aplicados a un amplio rango de dominios. Hay además normas en muchos campos para los cuales Dublin Core es de difícil aplicación. Ejemplso de estas normas incluye la norma de metadatos geoespaciales, las etiquetas meta del HTML y PRISM que es una norma de metadatos para la sindicación, agregación de contenidos de las revistas, catálogos y diarios digitales. Dublin Core está más orientado a la semántica de los metadatos que a la sintaxis y al lenguaje para codificarlo.

Qué lenguaje emplear para expresar metadatos. Este apartado ha sido objeto de atención especial en el ámbito de la edición tanto electrónica como en papel, y en Internet. Los últimos progresos han tenido lugar dentro de la comunidad de la Web Semántica a fin de poder codificar no sólo esquemas de metadatos sino también el modo en que se adptan entre sí para poder ser interoperables.

Normas acerca de las palabras a usar para los esquemas de metadatos. Hay una necesidad de normalización de las palabras empleadas para anotar datos y producir metadatos. Esta necesidad es patente y afecta desde la normalización de la ortografía a la normalización de la terminología. Esto hace que los vocabularios controlados se encuentren en constante proceso de rediseño, especialmente de aquellos lenguajes más adaptados al entorno digital.

Organizaciones supervisoras y promotoras de metadatos. La iniciativa de metadatos Dublin Core (DCMI) es la organización que trata con la definición, actualización y promoción de Dublin Core. Además las bibliotecas cuentan con autoridades, encabezamientos de materia y tesauros. Los metadatos bibliográficos fueron creados de forma cooperativa a partir de los años 60 y se difundieron a través de herramientas bibliográficas, OPAC's y bases de datos en línea.

4.2 Esquema Básico de Organización de la Información: Índice.

Antes de entrar en profundidad a ver en sí mismo el modelo de los topic maps, veremos cuál es el significado de las herramientas de ayuda a la navegación que nos son familiares en el ámbito de la Documentación, tales como las taxonomías, los índices de materia y los alfabéticos o bien los tesauros, las clasificaciones y las ontologías.

El concepto de índice fundamentalmente se relaciona con todo aquello que señala algo o que deja constancia de un indicio de algo³⁶⁶. En Documentación, no se habla tanto de índices como de lenguajes documentales, donde el índice sería una lista de palabras clave del lenguaje natural³⁶⁷. No obstante, los lenguajes documentales van más allá de una lista de palabras clave al insertarse en el ámbito de la terminología y los lenguajes de especialidad.

En cualquier caso, un índice y cualquier lenguaje documental, es un mapa del conocimiento acerca de un documento³⁶⁸, dado que enumera las materias cubiertas por el mismo y expresadas de manera clara para los usuarios que requieran su empleo. Así por ejemplo el índice del libro XML Topic maps³⁶⁹ nos dice: *Topic map Constraint Language (TMCL) development of, 340; user requirements for, 63-64 ; visualization of (see visualization); URIs(Uniform*

³⁶⁶ Diccionario de la RAE. En: www.rae.es Consultado el 20/12/2009

³⁶⁷ GIL URDICIÁIN, B. *Manual de lenguajes documentales*. Gijón : Trea, 2004, p.26.

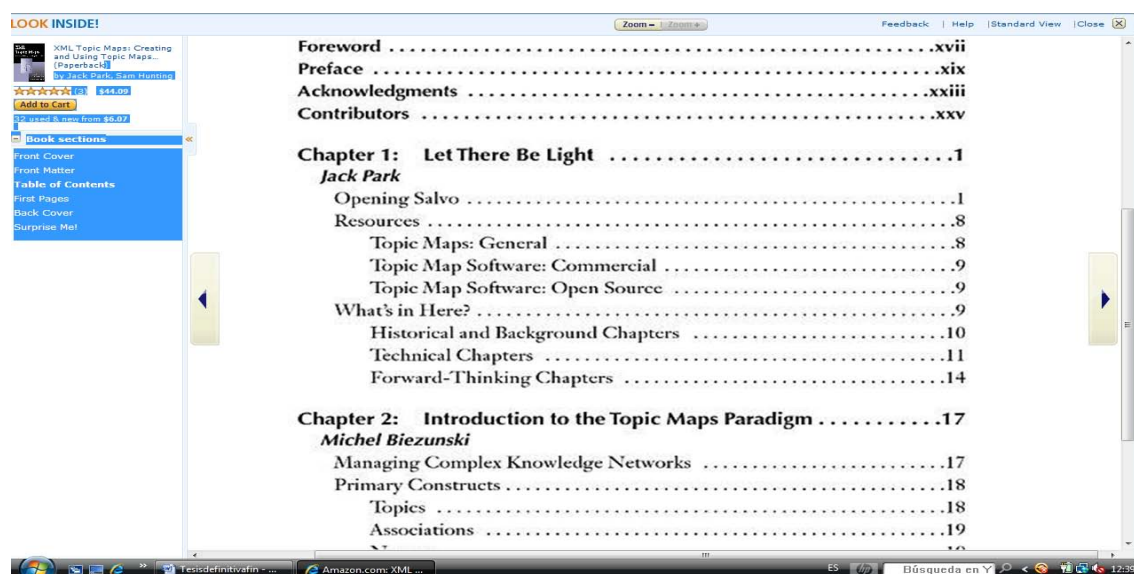
³⁶⁸ LANCASTER, F. *El control del vocabulario en la recuperación de información*. Valencia : Universitat de Valencia, 2002, p.31.

³⁶⁹ PARK, J. *XML Topic maps. Creating and using Topic maps for the Web*. Boston : Addison-Wesley, 2003. pp. 585ss

Resource Indicators) in Nexist, 259, 259f; XTM document and web pages transformation into, 13 (See also XML, and HTML).

Los principales elementos constituyentes de este y cualquier índice son:

- i) Una lista alfabética de nombres de materias.
- ii) Referencias a las páginas (occurrences) de las materias.



Foreword	xvii
Preface	xix
Acknowledgments	xxiii
Contributors	xxv
Chapter 1: Let There Be Light	1
Jack Park	
Opening Salvo	1
Resources	8
Topic Maps: General	8
Topic Map Software: Commercial	9
Topic Map Software: Open Source	9
What's in Here?	9
Historical and Background Chapters	10
Technical Chapters	11
Forward-Thinking Chapters	14
Chapter 2: Introduction to the Topic Maps Paradigm	17
Michel Biezunski	
Managing Complex Knowledge Networks	17
Primary Constructs	18
Topics	18
Associations	19

Figura 21. Ejemplo de sumario del libro de Jack Park. “XML Topic maps”.

Este ejemplo tomado del manual más conocido sobre Topic maps abarca materias(topics) muy diversas relacionadas con el tema, tales como la web semántica, la representación del conocimiento, la sintaxis XTM, o los lenguajes controlados. En un libro sobre otra materia, los topics y los tipos de topics sí que serían muy diferentes. En cuanto a los números de página, a donde nos remiten los topics, son parte de lo que denominamos en topic maps las occurrences. Sin embargo, existen otras referencias como los números de sección o los números de figuras o tablas que también forman parte de las occurrences. Todos ellos son localizadores de las materias en las fuentes de información.

En general, podemos decir que existen una serie de convenciones que permiten hablar de clases, tanto en topics como en occurrences. Las clases se establecen mediante el empleo de convenciones, que pueden ser³⁷⁰:

Tipográficas, empleadas para la distinción entre los distintos tipos de materias o de indicadores a las fuentes. En nuestro ejemplo podría ser el empleo de la cursiva para el encabezamiento de una familia de topics. Análogamente, las convenciones tipográficas son empleadas para diferenciar entre distintos tipos de occurrences o localizadores.

El empleo de subentradas, más conocidas en lenguajes documentales como referencias. Proporcionan un mecanismo alternativo para señalar asociaciones entre distintos topics. Entre ellas podemos destacar las referencias “véase” para tratar con la sinonimia que permite múltiples puntos de acceso a una materia, o “véase además” como punto de referencia para temas relacionados.

Además, los documentos pueden tener múltiples índices, como puedan ser los índices de nombres, de materias o de lugares, que no obstante, no se encuentran relacionados entre sí sino que los topics de varios índices de un mismo documento pueden llegar a apuntar recursivamente a idénticas secciones del mismo³⁷¹.

Los localizadores, que vienen a ser los números de las páginas en los índices, pueden contener modificadores que ayuden a distinguir entre distintos tipos de occurrences tales como “23n” para una nota de pie de página en la página 23 como alternativa al empleo de distintos tipos de letra. Otra manera de mostrar la naturaleza de una occurrence es mediante el empleo de mecanismos de subentradas³⁷² tales como: “definido en”, “definido en el glosario”.

Las características esenciales por tanto de un índice son los topics, identificados por sus nombres y que pueden tener más de uno, las associations entre los topics, y las occurrences marcados mediante los indicadores o url's asociados a cada topic. Y estos elementos

³⁷⁰ PEPPER, S. *The TAO of Topic Maps*. En: http://www.ontopia.net/topic_maps/materials/tao.html#d0e632. Consultado el 15/07/2009.

³⁷¹ Como veremos más adelante los topic maps proporcionan una forma alternativa al empleo de las convenciones tipográficas para la distinción de topics de distintos tipos en un mismo índice que a su vez reúne índices de muchas clases.

³⁷² GOLDFARB, Ch. ; PRESCOD, P. *Charles Goldfarb's XML handbook*. New Jersey : Prentice Hall, 2002, p. 33ss.

básicos veremos que también se encuentran entre clases de índices más complejos como los tesauros, las taxonomías y las redes semánticas.

4.3 Esquemas para la Organización Semántica de la Información.

El modo más sencillo para indizar documentos en una colección, y por tanto de proporcionales un acceso, es mediante la asignación de una colección de descriptores de un vocabulario controlado a un conjunto de documentos³⁷³. Este tipo de información se incluye en los documentos con un esquema de metadatos creado para dotar de un valor añadido al documento que facilite su organización, descripción, y acceso por parte de los usuarios³⁷⁴.

Los vocabularios controlados aseguran la uniformidad de la indización³⁷⁵, lo cual implica una elevada precisión en el momento de la recuperación. De hecho, asumiendo que el usuario sea capaz de encontrar el descriptor adecuado para describir su demanda informativa, entonces será capaz de encontrar todos los documentos descritos por el mismo. Por supuesto, la uniformidad de los descriptores es un hecho per se dado que el vocabulario es controlado. Las desventajas son que fuerza al usuario a acertar con el término de indización mas apropiado, y que los términos del índice son considerados como independientes, sin ninguna relación explícita entre sí. Estas dificultades se ven compensadas por la introducción de un amplio abanico de relaciones entre los descriptores de los vocabularios tales como las relaciones jerárquicas, de equivalencia y asociativas con su correspondiente abanico tipológico³⁷⁶. Identificaremos cada uno de los vocabularios controlados en función del tipo de relación que se pueda establecer entre los descriptores.

4.3.1 Taxonomías.

Las taxonomías son estructuras empleadas para la organización del conocimiento acerca de un particular dominio puesto que se refiere a la clasificación de objetos en categorías. Se

³⁷³ GIL LEIVA, I. *Manual de indización : teoría y práctica*. Gijón : Trea, 2008, p.37.

³⁷⁴ FOULONNEAU, M. *Metadata for Digital Resources : Implementation, Systems Design and Interoperability*. Oxford : Chandos, 2008. pp.15-16.

³⁷⁵ GIL URDICIÁIN, B. *Manual de lenguajes documentales*. Gijón : Trea, 2004. p.24

³⁷⁶ TUDHOPE, D. ; ALANI, H. ; JONES, C. Aumenting Thesurus Relationships: Possibilities for Retrieval. *Journal of Digital Information*, nº 1, febrero 2001. En: <http://jodi.ecs.soton.ac.uk/Article/v01/i08/Tudhope> Consultado el 27/12/09.

trata de un modo de categorizar una colección de objetos informativos, con la estructura de una jerarquía³⁷⁷.

En la norma ANSI/NISO Z39.19-2005 sobre construcción y uso de vocabularios controlados monolingües se define la taxonomía como³⁷⁸:

Una colección de términos de un vocabulario controlado organizada en una estructura jerárquica, donde cada uno de sus términos es parte, al menos de una, relación jerárquica genérica/específica que lo relaciona con otro término de la taxonomía.

Las taxonomías se caracterizan por un solo tipo de relación, la jerárquica. Este tipo de relación muestra niveles de sobreordenación y subordinación. El término sobreordenado representa una clase y los términos subordinados representan los miembros de esas clases. Esta relación es empleada para la localización de conceptos más genéricos o específicos de acuerdo con una secuencia progresiva lógica. Se trata de una característica básica de las taxonomías y un importante factor para la mejora de la exhaustividad y la precisión en la recuperación por parte de los motores de búsqueda.

La relación jerárquica incluye la relación genérica por la cual los términos se relacionan jerárquicamente sólo si los miembros de la misma categoría fundamental representan al mismo tiempo entidades, actividades, agentes o propiedades. La relación genérica es definida como aquella por la cual se identifica el enlace entre una clase o una categoría y sus miembros o especies³⁷⁹.

A una taxonomía se la describe con raíz como primer elemento superior, donde cada nodo es una entidad de información que se corresponde con una entidad del mundo real. Cada uno de los enlaces entre nodos representa bien una relación del tipo “es subclasificación de” o bien una relación tipo “es superclasificación de”. En ocasiones a estas relaciones

³⁷⁷ FENSEL, D. *Spinning the Semantic Web: bringing the World Wide Web to its full potential*. Cambridge : MIT press, 2003, p.146.

³⁷⁸ *Guidelines for the Construction, Format and Management of Monolingual Controlled Vocabularies*. En: http://www.niso.org/kst/reports/standards/kfile_download?id%3Astring%3Aiso-8859-1=Z39-19-2005.pdf&pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVa86hjP37r_hONsJghRDv2N-zj4TZCh8Dp01rZbmK3O-8vcVjh4hezP.

Consultado el 20/12/09.

³⁷⁹ AITCHINSON, J. ; GILCHRIST, A. *Thesaurus construction and use: a practical manual*. Londres : Aslib, 1990. pp.39-40.

especiales se las define de manera más estricta como “es subclase de” o es “superclase de”, donde se entiende que las entidades de información son clases de objetos³⁸⁰.

A medida que se navega hacia la raíz de la taxonomía, las entidades se van haciendo genéricas, y contrariamente al descender se van haciendo más especializadas. Es por ello que a este tipo de sistema de clasificación se le denomina taxonomía de generalización-especialización o de agrupación³⁸¹: un ejemplo conocido es la taxonomía de especies de Linneo. Una taxonomía puede también emplear la relación de parte-todo en cuyo caso se la conoce por meronimia.

Una taxonomía, desde el punto de vista de la organización del conocimiento, es una jerarquía semántica en la cual las entidades de información se encuentran interrelacionadas por la relación “subclasificación de” o bien “subclase de”, siendo la primera relación más débil que la última. Esto permite distinguir entre taxonomías fuertes y débiles, si bien las taxonomías tienen en sí mismas una semántica débil, según empleen la propiedad o atributo de diferenciación. Cada una de las entidades de información es diferenciable por su propiedad de diferenciación que las hace únicas como subclase de su entidad superior. Normalmente la propiedad que caracteriza una subclase específica en el nivel más alto es una colección de propiedades o atributos.

Un ejemplo de ello lo tenemos si consideramos una taxonomía para ciencias de la información³⁸². La distinción entre la clase “2.1 Thesauri, authority lists” y la clase “2.2 Cataloging and classification” bajo la superclase “2. Knowledge Organization”. Aunque tanto las listas de autoridades como la clasificación tienen la propiedad común de ser una lista de términos, las listas de autoridades no tienen relaciones de jerarquía en tanto que las clasificaciones si las tienen. Por tanto el hecho de poseer la propiedad de contar con la relación de jerarquía constituye una propiedad diferenciadora entre ambas clases, pero existen otras propiedades que también las diferencian como puede ser la propiedad de tener la relación parte/todo.

³⁸⁰ JAGERMAN, E. *Creating, maintaining and applying quality taxonomies*. Zoetermeer, 2006. p.28.

³⁸¹ En la literatura inglesa sobre taxonomías se le denomina subsumption que es una palabra latina derivada de subsumere que significa ser parte de algo más general.

³⁸² HAWKINS, D.T. ; LARSON, B. Q. Information Science Abstracts: Tracking the Literature of Information Science. Part 2: A new taxonomy for information science. *Journal of the American Society for Information Science and Technology*, vol. 54, nº8, 2003, pp.771-781.

Sin embargo, en la taxonomía cuanto más bajo es el nivel, más probable es que una sola propiedad distinga una subclase específica. En general se puede decir que, si la taxonomía es semánticamente fuerte y con una estructura bien definida, hay una propiedad diferenciable y específica para cada clase de la taxonomía. Es más, la especificidad o granularidad de la propiedad se incrementa a medida que se desciende en la taxonomía.

El ámbito de las taxonomías y de las estructuras taxonómicas está muy ligado a distintas corrientes filosóficas, pero debido a la necesidad de organizar la información en el ámbito digital ha obligado a ampliar su significado para aplicarlo en la organización del conocimiento en un contexto donde el significado de documento se ha ampliado. En este contexto se usa actualmente las taxonomías para la generación de metadatos, la recuperación de información, la generación de categorías que asistan al usuario durante la navegación a través de un repositorio documental, la generación de esquemas que rijan la configuración y estructura de los sitios web, y la creación de listas controladas de términos que facilitan labores de web mining o localización de información.

Hoy en día existen muchos organismos implicados en proveer a sus usuarios de acceso a información semiestructurada e incluso a información desestructurada, donde la integración con otros esquemas de información y su acceso resulta difícil. En este sentido las taxonomías vienen a cumplir un papel de enlace entre los autores de los documentos y los contenidos o banco de conocimientos que cada organismo quiere difundir, bien internamente o bien de forma universal³⁸³.

El hecho de poder asignar nuevos significados a los términos de un dominio de conocimiento ya establecido debe hacernos pensar en la necesidad de ser capaces de adaptar los lenguajes controlados tradicionales a los cambios en el discurso recogido en los documentos. Y uno de los primeros lenguajes controlados afectados por este hecho son las taxonomías, que han sufrido en los últimos años un replanteamiento de sus posibilidades.

Uno de los autores que más se ha planteado el nuevo papel de las taxonomías, y más concretamente de las taxonomías corporativas en el mundo de la documentación digital y la

³⁸³ VOGEL, C. *Quality metrics: how to ensure quality taxonomies*. San Mateo: Semio Corporation, 1999.

organización del conocimiento ha sido Gilchrist, quien considera que una taxonomía (corporativa) aspira a ser³⁸⁴:

...una correlación de los diferentes lenguajes funcionales, regionales y nacionales utilizados por la organización para dar soporte a un mecanismo de navegación y acceso al capital intelectual de la organización...

Y apunta los siguientes factores como catalizadores en el cambio de concepción de las taxonomías³⁸⁵:

- a) la sobrecarga de información. Los motores de búsqueda convencionales se han manifestado en numerosas ocasiones incapaces para tratar de manera efectiva con grandes bases de datos sin contar con filtros y asistentes para las búsquedas, especialmente para los usuarios,
- b) La escasa alfabetización informacional de los usuarios. Las investigaciones acerca de los hábitos de consulta de información por parte de los usuarios han revelado importantes carencias en habilidades para la búsqueda de información, evidenciando un elevado gasto de tiempo para la poca efectiva localización de información adecuada para las demandas de información,
- c) La mala adecuación del vocabulario de los lenguajes controlados al vocabulario de los sitios web u organizaciones en los que se aplican, particularmente en dominios con un lenguaje muy especializado como es el derecho,
- d) Desestructuración de la información en la red, dado que la distribución de información por la red ha traído problemas en la interoperabilidad de documentos de un mismo dominio de conocimiento por ser generados en ámbitos culturales diferentes³⁸⁶. Esto es un obstáculo para la constitución de comunidades virtuales que comparta un mismo lenguaje

³⁸⁴ GILCRIST, A. ; KIBBY, P. ; MAHON, B. *Taxonomies for business: access and connectivity in a wired world*. London : TFPL, 2000, p.6

³⁸⁵ GILCHRIST, A. Thesauri, taxonomies and ontologies- an etymological note. *Journal of Documentation*, vol.59, n°1, 2003. pp.7-18.

³⁸⁶ Pensemos por ejemplo en los problemas de recuperación de información de documentación jurídica del ámbito anglosajón con una tradición jurídica completamente diferente a la española, la italiana o la francesa.

controlado y más aún si tienen que crearlo para la recuperación e indización de información en un sitio web multilingüe.

Cuando se pretende afrontar todos estos problemas de las taxonomías en el entorno digital se evidencia que el concepto de taxonomía es amplio y genérico, aglutinando un amplio rango de técnicas y aplicaciones. Para Gilchrist la palabra taxonomía puede ser empleada hasta con cinco significados diferentes, si bien algunos de ellos se solapan parcialmente³⁸⁷.

Por un lado una taxonomía puede ser entendida como un directorio a ser empleado en los sitios web o en las intranets de las organizaciones, de hecho se trata de una forma genérica de clasificación a partir de una lista alfabética de topterms, que se le ofrecen a cualquier usuario que acceda al sitio. Por medio de hipervínculos se posibilita la navegación, pero no estrictamente tiene por qué ser en un orden jerárquico. La navegación se termina hasta que el usuario localice el recurso que satisfaga la demanda informativa del usuario. Por otra parte, los términos que constituyen la taxonomía pueden aparecer repetidos en distintos niveles de la misma, lo cual facilita el acceso a un mismo recurso mediante distintas estrategias de navegación.

Otro sentido para las taxonomías es como soporte para la indización automática. Nos referimos con esto al caso de los sitios web comerciales que presentan una clasificación a dos niveles para sus productos: i) la clasificación que se le presenta al usuario en la interfaz y, ii) la taxonomía que está detrás de esa misma clasificación y que el usuario no ve. En esta taxonomía se contienen los algoritmos que permiten relacionar para cada término mostrado en la clasificación una serie de palabras y frases con sus variantes sintácticas y sinonímicas así como sus ponderaciones. Estos algoritmos son las reglas básicas por las cuales es posible una adecuada extracción automática de términos de indización con independencia de que se encuentren o no en los documentos que se indizan. Esta estrategia es la más adecuada para grandes repositorios de información en los cuales la indización manual resulta económicamente poco viable. No obstante tanto la clasificación como la taxonomía son construidas de forma manual, y la taxonomía contiene tantas relaciones que más bien puede ser considerado como un tesoro bastante desarrollado.

³⁸⁷ GILCHRIST, A. Op. cit., p.11.

Un tercer significado es para aquellas taxonomías creadas a partir de la definición automática de clases. Un número creciente de programas de software son capaces de analizar textos en lenguaje natural, crear automáticamente categorías a partir del análisis y clasificación de los documentos analizados de acuerdo con unas categorías previamente definidas. Estas categorías pueden ser mostradas en un formato análogo a los directorios web o bien mediante mapas bidimensionales, donde los términos relacionados están enlazados a términos seleccionados que aparecen en el medio del mapa. Al seleccionar un término relacionado, éste se desplazará al centro de la pantalla como un centroide y a su vez se dispone a su alrededor una nueva colección de términos relacionados³⁸⁸. Como en el caso del directorio web, el resultado del proceso de categorización automática es una clasificación, y a menudo el empleo de un software de recuperación probabilística se basa en el análisis estadístico de la ocurrencia y coocurrencia de términos en la base de datos, aunque aumentados por la intervención humana.

Otro significado que se le confiere es como filtro front-end. Aquí una taxonomía es o bien creada o bien importada y usada en la formulación de consultas. Los términos homógrafos pueden ser desambiguados y los sinónimos agrupados en cluster; mientras que el usuario puede navegar jerárquicamente entre los términos y saltar a otros relacionados. Una vez seleccionado el término de búsqueda es enviado a Internet o a la intranet. Este tipo de taxonomía no es más que un tesoro formateado para permitir una fácil navegación.

Finalmente puede asignársele un significado de taxonomía corporativa. Los ejemplos precedentes de distintas taxonomías son empleados tanto a nivel de procesamiento en la entrada de información o bien para mostrar los términos de búsqueda y los resultados. Sin embargo, las empresas actuales se ven ante importantes problemas de sobrecarga de información, así como dificultades en la interoperabilidad de sistemas de gestión de contenidos, para lo cual los vendedores de programas de gestión documental están proporcionando soluciones a través de la inserción de taxonomías de asistencia en la búsqueda al usuario. La necesidad primordial de hacer que la información sea fácilmente accesible a su personal a través de un portal de información para la empresa u otro tipo de canal supone hablar al mismo tiempo de grandes bases de datos y de una gran capacidad de procesamiento. Hay, además, una necesidad por proporcionar mapas y guías a unos repositorios documentales que pueden contener una ingente cantidad de documentos, de

³⁸⁸ Véase el programa *Thinkmap Visual Thesaurus*. En: <http://www.visualthesaurus.com/>. Consultado el 15/12/2009.

los que parte de ellos pueden encontrarse fuera del repositorio. Esto puede llevar a un empleo erróneo del concepto de mapeo, en el sentido de implicar la necesidad de emplear un gran esfuerzo intelectual.

Un ejemplo de este tipo de taxonomía es la aplicada en la empresa Glaxo Smithkline³⁸⁹, donde una serie de tesauros han sido unidos libremente en un megatesauro; en el cual, cada término está hiperenlazado a un repositorio de información corporativo que emplea ese mismo término. Además, esos mismos términos contienen información de autoridad y derechos de uso sobre ellos. Este gran lenguaje documental que incluye 53500 descriptores, más de doscientos mil sinónimos y casi medio millón de términos relacionados, de los cuales casi todos tienen bien delimitada el tipo de relación asociativa asignada. Esta empresa se ha aliado con varias empresas de software especializado en ontologías para el desarrollo y venta de este tesoro ahora convertido en una ontología capaz de asimilar nuevos tesauros sin que suponga una merma en su grado de precisión en la recuperación. En este sentido, aunque podría ser considerado como un sistema híbrido de tesoro/ontología, se trata de un nuevo tipo de taxonomía en emergencia denominada “taxonomía corporativa”.

Este tipo de taxonomía, necesita ser construida manualmente, proporciona un mapa de alto nivel para un portal de empresa, guiando al personal en la navegación por el capital intelectual de la organización. Puede abarcar desde información acerca de la propia empresa como guías, normas o métodos hasta información que se encuentra en repositorios ajenos a la propia empresa. Además puede ser diseñada para asistir a los motores de búsqueda, o bien para proporcionar plantillas al indizador. Esta taxonomía puede ser vista como un esquema maestro a partir del cual generar otros lenguajes documentales como tesauros, topic maps u ontologías.

Al tratarse de la taxonomía corporativa de una empresa no podemos recoger un ejemplo de la misma, pero sí en cambio contamos con una taxonomía corporativa parecida con la que cuenta la Organización Mundial de la Salud (OMS). Se trata de la taxonomía del “Centers for disease control and prevention³⁹⁰”. Como vemos en la primera figura, contamos con

³⁸⁹ Se trata del proyecto Babylon para la creación de una plataforma de text-mining. En: <http://ieeexplore.ieee.org/iel5/6294/26712/01191789.pdf>. Consultado el 25/07/2009

³⁹⁰ Taxonomía corporativa del Centers for disease control and prevention. En: <http://www.cdc.gov/az/m.html>. Consultado el 15/07/2009.

Topic Maps y Tecnologías de la Información Jurídica: un modelo de recuperación de información para bibliotecas digitales semánticas jurídicas.

una relación jerarquizada de topics hiperenlazados a sitios web del propio centro que contienen todo tipo de información sobre el topic.

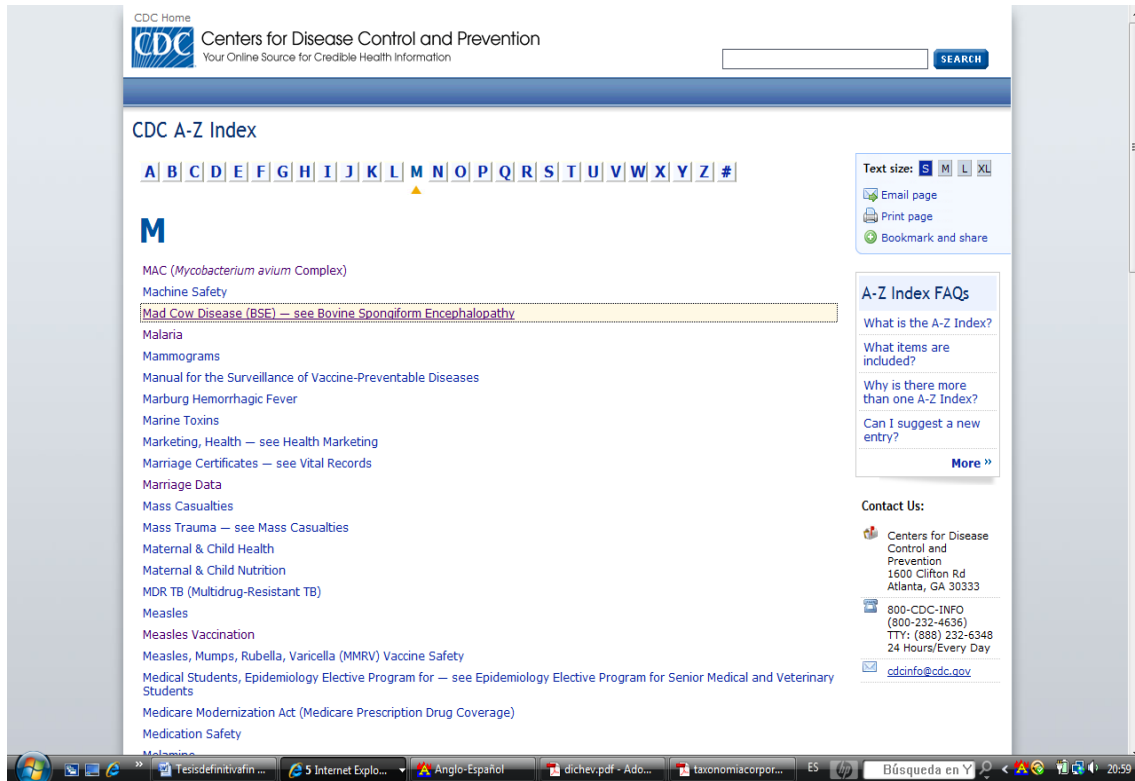


Figura 22. Ejemplo de taxonomía del Centers for Disease Control and Prevention.

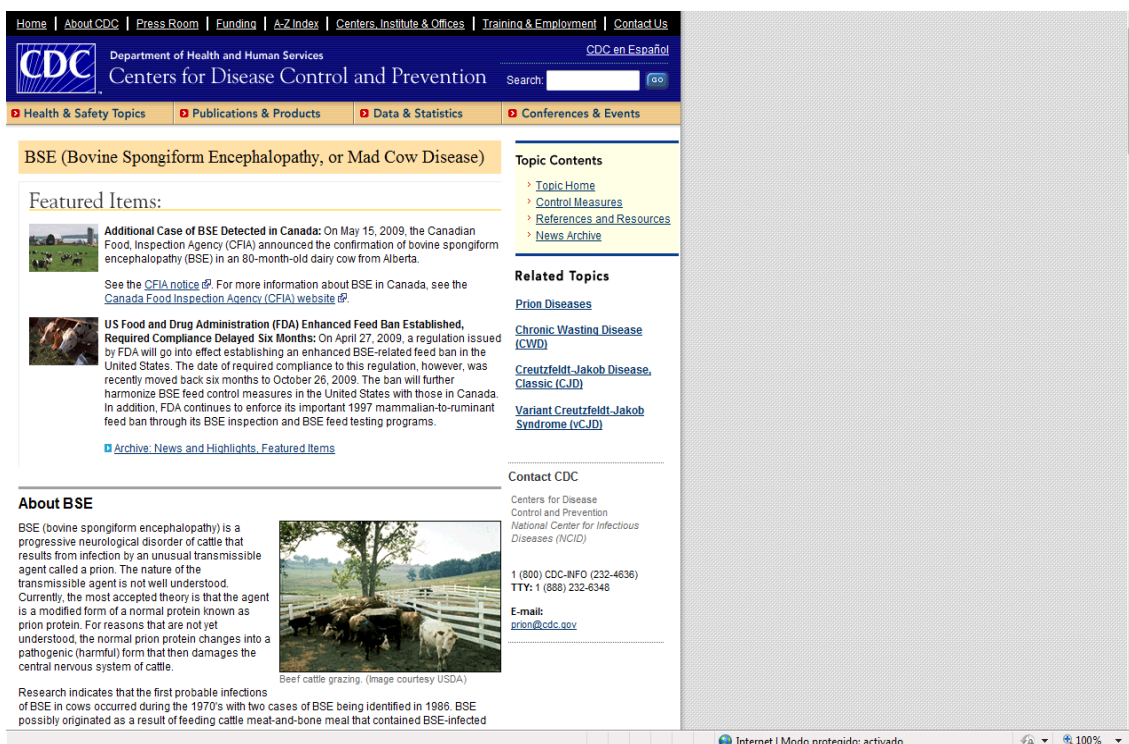


Figura 23. Ejemplo de hiperenlace desde el topic “Encefalopatía espongiforme” al sitio web del centro sobre la enfermedad.

4.3.1.1 Metodología en la construcción de taxonomías.

La creación de taxonomías es, para Jagerman, en parte una ciencia y en parte un arte³⁹¹. En su nivel más elevado supone una sutil comprensión de la naturaleza humana y de su entorno cultural, así como una buena gestión de las necesidades de los usuarios y los objetivos del centro.

En el momento de construir la taxonomía resulta útil pensar en los siguientes factores:

- a) a quién va destinada la taxonomía y qué cantidad y tipo de información requiere para la resolución de sus demandas informativas,
- b) simplicidad y pragmatismo en su empleo,
- c) Fácil mantenimiento de forma que a los usuarios les resulte fácil la navegación dado que un esquema de clasificación amplio facilita considerablemente la navegación,

³⁹¹ JAGERMAN, E. *Creating, maintaining and applying quality taxonomies*. Zoetermeer, 2006, p. 37.

- d) decidir la profundidad que se desee proporcionar al usuario durante la navegación en la taxonomía así como conocer el nivel de aceptación,
- e) permitir la evolución y crecimiento orgánico durante la construcción, lo cual mejorará las posibilidades de éxito,
- f) conocer el tiempo y dinero que se puede invertir en el proceso de creación,
- g) averiguar el nivel de aceptación dentro de la organización,
- h) gestionar las expectativas de los usuarios en cuanto a la localización de los recursos de información.

Antes de crear una taxonomía es aconsejable hacer una investigación previa de taxonomías ya existentes sobre la misma temática. La reutilización de taxonomías, al igual que la reutilización de tesauros³⁹² puede ayudar a ahorrar mucho tiempo y esfuerzo. Existen ya establecidas listas de autoridad y lenguajes controlados sobre multitud de temáticas que en muchas ocasiones son accesibles gratuitamente en línea³⁹³ o bien en bibliotecas tanto públicas como especializadas o en centros de documentación especializados. En cualquiera de estos casos lo primero que se debe hacer es averiguar si el acceso a estos lenguajes documentales es gratuito o no.

Como punto de partida para la elaboración de un vocabulario básico de búsqueda se debe examinar cómo se refleja el comportamiento del usuario en las sentencias de búsqueda. Estas son el mejor indicio para averiguar el tipo de información que los usuarios suelen emplear, los términos que usan e incluso los errores más frecuentes en la formulación de las consultas. Con esta investigación se busca determinar la frecuencia de términos más empleados³⁹⁴.

³⁹² Véase la tesis sobre el empleo de tesauros para la reutilización de software de LLORENS, J. *Definición de una metodología y una estructura de repositorio orientadas a la reutilización : el tesoro de software*. Leganés, 1996.

³⁹³ Véase la lista de tesauros en línea del CINDOC. En: www.cindoc.csic.es. Consultado el 15/12/2009.

³⁹⁴ De acuerdo con la Ley de Zipf. GIL LEIVA, I. *Manual de indización. Teoría y práctica*. Gijón : Trea, 2008. p.334.

Otra fuente de información para el documentalista es la retroalimentación que aportan los informes de los archivos logs³⁹⁵ con información acerca de los hábitos de consulta en sitios web para la evaluación de su arquitectura de información. De acuerdo con estos archivos se comprueba que la habilidad en el manejo del lenguaje de especialidad del dominio en el que se hace la búsqueda es esencial para el acceso a la información dado que en muchas ocasiones los usuarios no son capaces de localizar la información deseada, a pesar de contar con herramientas de asistencia como tesauros o clasificaciones, bien por su desconocimiento del lenguaje especializado o bien porque los documentos han sido mal indizados por parte de unos documentalistas desconocedores del lenguaje de especialidad³⁹⁶. A los informes de hábitos de búsqueda proporcionados por los logs se le pueden añadir los informes realizados con las técnicas del *focus group*, las entrevistas contextuales y los informes de usabilidad.

Por medio de encuestas abiertas a los bibliotecarios referencistas es posible conocer qué recursos de información son los más consultados y al mismo tiempo los más difíciles de localizar, lo cual permite adaptar la taxonomía del centro a las demandas informativas del usuario y desarrollar los términos de la taxonomía de forma que los más utilizados o bien los más adecuados pero no tan utilizados sean usables al usuario.

Una taxonomía bien constituida no sólo refleja las necesidades del documentalista en relación con su labor de hacer accesible su colección o repositorio documental, sino que también refleja el contenido organizado. De hecho un método de construcción de taxonomías según el contenido es a partir de la determinación del uso actual de los metadatos a partir de las etiquetas meta insertadas en las cabeceras de los documentos digitales.

La construcción de taxonomías a partir de los metadatos de los documentos parte de una colección básica y común de elementos meta muy utilizados en los documentos de nuestro repositorio, de forma que aunque los documentos que contenían esas etiquetas meta ya no existan en el repositorio se conservan a causa de los *schemas* con que han sido construidos

³⁹⁵ SERRANO-COBO, J. Combinación de logs internos y externos en la predicción de estacionalidad de búsquedas para el rediseño de webs. *El Profesional de la Información*, vol.18, nº1, enero-febrero 2009.

³⁹⁶ Realidad especialmente palpable en centros de documentación muy especializados como en bibliotecas hospitalarias o jurídicas.

los sitios web que contienen esos documentos, aunque en la práctica esas etiquetas meta no hayan sido empleadas ni se les haya asignado ningún valor.

También se debe tener en cuenta para la construcción de una taxonomía que tanto los organismos propietarios de los documentos como sus usuarios proporcionan una visión global de las materias tratadas en su colección y de si éstas son adecuadas para la localización correcta de la información que se pretende difundir. En este sentido cobra interés tener en cuenta las opiniones de los responsables de los repositorios para saber si los documentos que contienen están siendo recuperados correctamente por sus potenciales usuarios o no. Y más si tenemos en cuenta que los responsables de los repositorios y sitios web pueden ser expertos en las materias de los documentos a distribuir, los cuales nos pueden indicar mejor las fortalezas y debilidades de la localización de contenidos en esos dominios de conocimiento.

Los autores han llegado al acuerdo de que fundamentalmente existen tres métodos³⁹⁷ para la construcción de taxonomías, si bien señalan que en la práctica no se aplican cada uno de ellos de forma separada sino que en muchas ocasiones existe una combinación entre ellos³⁹⁸.

El método *top-down* (de arriba abajo). Es el método más tradicional de creación de esquemas de clasificación. Por medio de esta aproximación los expertos de un dominio analizan las materias y determinan con claridad donde situar los documentos en un listado de categorías previamente definido. El modo de trabajar se muestra en la siguiente tabla.

Paso	Tarea
Recoger y registrar el material	Selección de recursos de información para la constitución de una colección de términos. Esta tarea se efectúa por medio de entrevistas a los especialistas de un dominio haciéndoles examinar una colección de fuentes de información, sentencias de búsqueda y posibles términos para la indización.

³⁹⁷ JAGERMAN, E. *Creating, maintaining and applying quality taxonomies*. Zoetermeer, 2006. pp.42-47.

³⁹⁸ HAMMOND, R. The science of classification. *Professional Webmaster*, july/august 2000. p.20.

	<p>Asignar un código abreviado a cada uno de los entrevistadores para conocer de quién ha planteado cada término.</p> <p>Hacer una selección de términos y transferir los términos seleccionados con toda la información (definiciones, sinónimos y términos relacionados) en forma taxonómica al ordenador.</p>
Lematización de todos los términos en orden alfabético y unificación de toda la información sobre términos idénticos en una única forma.	Lematización de todos los términos preferentes en un orden alfabético. Unificación de información para términos idénticos e incluso puede que sea necesario añadir información adicional de otras fuentes para aclarar las razones de algunas unificaciones de terminológicas.
Elaboración de una estructura sistemática preliminar de la taxonomía como puede ser la creación de un mapa conceptual de todos los términos. Y adición de técnicas tesaurii como la inserción de estructuras de homonimia-sinonimia, o de equivalencia.	<p>Trabajo detallado de la estructura. Definición amplia de las materias e inserción de los términos en cada uno de estos campos.</p> <p>Posteriormente añadir subcampos a los precedentes campos genéricos y ordenar los términos en cada uno de los subcampos.</p>
Elaboración de un primer borrador del esquema.	Escribir una primera versión de la taxonomía y distribuirla entre los expertos del dominio de la taxonomía para que éstos hagan sugerencias acerca de la estructura y poder posteriormente efectuar mejoras.
Completar la taxonomía ya con todos sus términos.	<p>Repaso de todas las entradas en un archivo de trabajo como sigue:</p> <p>Formulación normalizada de las abreviaturas.</p> <p>Establecimiento de los descriptores y de los términos relacionados y jerárquicos. Mejora de la estructura clasificatoria definiendo la relación de los términos con sus descriptores.</p>
Comprobar el funcionamiento de la taxonomía por medio de experimentos de indización y recuperación con la misma.	Evaluación de la taxonomía por medio de ensayos en un corpus documental del dominio y hacer las modificaciones oportunas.

Publicación de una versión para usuario de la taxonomía.	Aplicación de la taxonomía en la indización y recuperación.
--	---

Tabla 5. Método *top-down*. Elaboración propia.

El método *bottom up* (de abajo arriba). Se parte de las propiedades agrupadas en clases. Se comienza por separar una subcolección de un repositorio de documentos de acuerdo con una categoría específica. En caso de que en el repositorio haya documentos susceptibles de ser situados en categorías diferentes se crean nuevas subcategorías hasta poder situar todos los documentos en una única subcategoría. Por tanto, este método de construcción de taxonomías se basa en la subdivisión repetitiva de categorías como parte del proceso clasificatorio³⁹⁹.

Paso	Tarea
Constitución de un grupo de expertos en el dominio del que se quiere crear la taxonomía.	Trabajo con los expertos para determinar el alcance de la taxonomía.
Recoger una colección representativa de documentos ya indizados.	Empleo de los términos de indización, obtenidos a partir de un repositorio documental, como lista preliminar de términos
Si los documentos no están indizados, crear un repositorio de documentos.	Creación de una lista de términos a través de la indización manual y libre de un repositorio de documentos. Además también es posible el empleo de la lista de términos del índice que pueda haber del repositorio documental.
Creación de una primera versión del esquema taxonómico.	Construcción de una taxonomía por revisión y organización de sus términos.
Contactar con los expertos en la materia	Acudir a los expertos en el dominio de la temática de la taxonomía cuando no esté claro el

³⁹⁹ Delphi Group. *Taxonomy and content classification*, 2002. En:

http://lsdis.cs.uga.edu/SemanticEnterprise/Delphi_LingoMotorfinal.pdf . Consultado el 15/05/2009

	significado de los términos.
Evaluación de la taxonomía por medio de ensayos de indización y recuperación.	Evaluación de la taxonomía por medio de sucesivos experimentos de indización y recuperación de colección de documentos. Inserción de las necesarias modificaciones.
Publicación de la versión final de la taxonomía.	Aplicación de la taxonomía a la indización y recuperación.

Tabla 6. Método *bottom-up*. Elaboración propia.

La aproximación facetada. Muchos lenguajes documentales clasifican los términos de sus vocabularios en categorías muy genéricas denominadas facetas. Las facetas son agrupaciones de términos obtenidas por medio de una primera división de una materia en categorías semánticas coherentes y homogéneas. Que una faceta sea semánticamente coherente, significa que sus términos se encuentran relacionados entre sí por medio de relaciones paradigmáticas de sinonimia y jerarquía, y que la totalidad de facetas utilizadas en un lenguaje documental son mutuamente excluyentes⁴⁰⁰.

El análisis de facetas implica⁴⁰¹:

- a) la creación de una colección de términos que representan simples conceptos,
- b) la agrupación de términos bajo categorías mutuamente excluyentes denominadas facetas, empleando tan sólo una característica distintiva a la vez,
- c) organización de las facetas en un número limitado de categorías fundamentales. Estas son categorías genéricas que pueden ser adaptadas a cualquier materia y definen la función de un término dentro del esquema global del tesaurus,
- d) en muchos casos se necesitará de una notación para la fijación del valor clasificatorio de cada término en una secuencia sistemática.

⁴⁰⁰ AITCHISON, J. *Thesaurus Construction and use: a practical manual*. London: Aslib, 1990.

⁴⁰¹ ROWLEY, J. *Organizing knowledge: an introduction to managing access to information*. Aldershot: Gower Publishing Limited, 2000.

La partición de un vocabulario controlado en facetas es muy útil cuando se trata de crear agrupaciones de términos que puedan ser más o menos manejables. Los términos pueden ser agrupados uno a uno en grupos que limitan el dominio de los mismos. Esto simplifica y al mismo tiempo asegura la corrección en la estructuración de las relaciones jerárquicas y sinonímicas tal y como vemos en el siguiente ejemplo para el término MEDLINE en la taxonomía de MeSH.

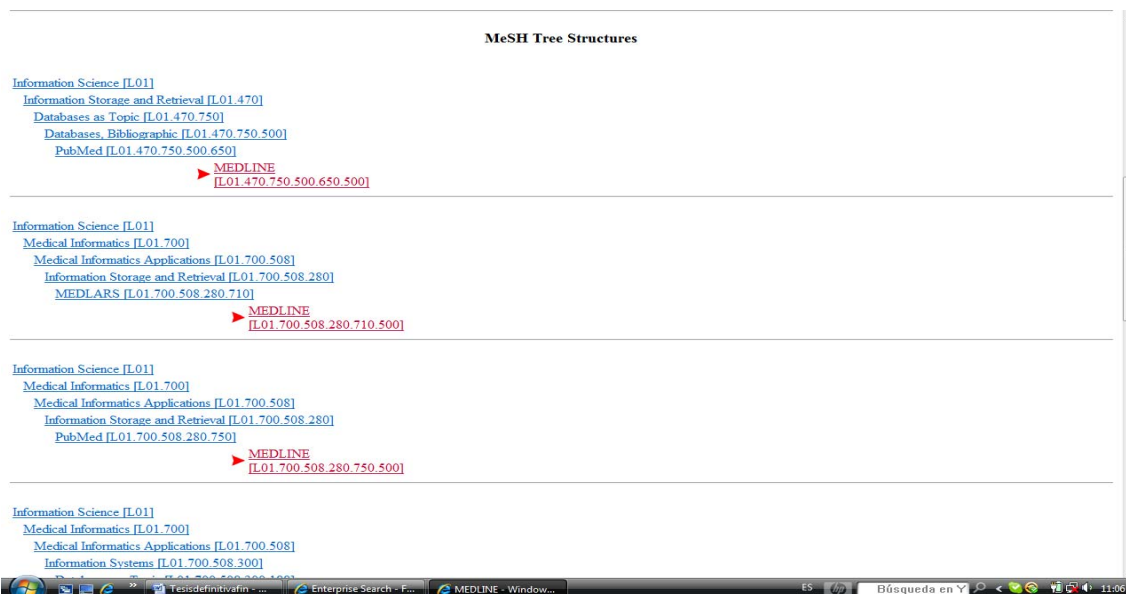


Figura 24. Ejemplo de 6 formas de acceso a la material MEDLINE en la taxonomía de MeSH.

El mayor problema de las facetas es que resultan difíciles de definir y por tanto dificulta la posibilidad de integrar los términos en clases. Estas dificultades son causadas por el hecho de que el lenguaje libre es a menudo poco estructurado. Así hay términos que no pueden ser categorizados exclusivamente en una única faceta, y términos difíciles de categorizar por su ambigüedad o su elevada abstracción⁴⁰².

4.3.2 Tesoros.

El término tesoro tiene su origen etimológico en la palabra latina thesaurus, que, a su vez, tiene su origen en la palabra griega thesaurós. En ambos casos, el significado era tesoro o repositorio de palabras⁴⁰³. La lexicografía fue el primer campo en el que se aplicó el término

⁴⁰² SVENONIUS, E. *The intellectual foundation of information organization*. Cambridge: MIT Press, 2000. pp.165ss.

⁴⁰³ CURRÁS, E. *Ontologías, taxonomías y tesauros*. Gijón : Trea, 2005. pp.78ss.

tesauro y en el que se elaboraron las primeras definiciones. Así Roget⁴⁰⁴, definió el tesauro como una colección de palabras y frases ordenadas, no en orden alfabético, como están en un diccionario, sino de acuerdo con las ideas que representan. Es decir, que se tiene la idea, y hay que buscar la palabra, o palabras que se ajusten más exactamente a esa idea. Es decir, su finalidad es la de proporcionar ayuda y asistencia a la expresión de ideas. Por tanto tiene una organización conceptual, donde a partir de los conceptos explicitados en las entradas se relacionan y agrupan las palabras que designan o nombran, en diversos contextos el concepto en cuestión.

A partir del desarrollo de los sistemas de indización y clasificación sobre la base de palabras clave Howerton concibe el tesauro como una lista autorizada que puede conducir al usuario de un concepto a otro por medio de relaciones heurísticas o intuitivas⁴⁰⁵. Lista que se puede usar manual o mecánicamente para la asignación de encabezamientos de indización.

Vickery expuso cuatro posibles significados para el término thesaurus en Documentación, de entre los cuales la definición mas aceptada ha sido aquella que entiende por tesauro una lista alfabética de palabras, donde cada palabra está desarrollada con otra lista de palabras relacionadas⁴⁰⁶.

Gilchrist apunta que se usa por primera vez el término tesauro en el campo de la Documentación⁴⁰⁷, en un trabajo presentado a la Dorking Conference on Classification por Helen Brownson, miembro de la American National Science Foundation. En dicha contribución, el término tesauro es utilizado al analizar la problemática de traducir los conceptos y sus relaciones, tal como se expresan en los documentos, a un lenguaje con mayor precisión y sin ambigüedades, con los sinónimos controlados y sus estructuras sintácticas simplificadas para facilitar la recuperación de información.

Wersig afirma en su obra que un tesauro es una lista de términos, prefijados con anterioridad, pero sacados del texto de los documentos, que desdobra los conceptos en las

⁴⁰⁴ HÜLLEN, W. A History of Roget's Thesaurus. Origins, Development and Design. Oxford : Oxford University Press, 2004. pp.327ss.

⁴⁰⁵ LÓPEZ HUERTAS, M.J. Thesaurus structure design: a conceptual approach for improved interaction. *Journal of Documentation*, vol. 53, nº 2, 1997. pp.159-167.

⁴⁰⁶ FOSKETT, D. *Subject and Information analysis*. New York : Marcel Dekker, 1985. pp.270-316.

⁴⁰⁷ GILCHRIST, A. Classification and thesauri. En VICKERY, B.C. (ed.). *Fifty years of information progress: a Journal of Documentation Review*. London: Aslib, 1994. pp.85-118.

unidades simples. Estas se coordinan posteriormente para evitar ambigüedades⁴⁰⁸. Entre ellas se establecen relaciones jerárquicas, asociativas y de equivalencia. De acuerdo con esto, Gilchrist afirma que un tesoro es una lista autorizada de léxicos, sin notación, que difiere de una lista de encabezamientos de materias, en las que unidades léxicas, siendo más pequeñas, son más manejables y se utilizan en la indización coordinada. Si bien en una edición posterior de su obra, al adaptarla a la recuperación de información digital afirma que el tesoro es un vocabulario de un lenguaje de indización controlado, de forma que:

“a priori” se hacen explícitas las relaciones entre conceptos, para ser usadas en los sistemas de recuperación de información, saltando desde el catálogo de fichas a Internet⁴⁰⁹.

En el contexto normativo, AENOR (Asociación Española de Normalización y Certificación) define un tesoro en los siguientes términos:

Vocabulario de un lenguaje de indización controlado organizado formalmente con objeto de hacer explícitas las relaciones, a priori, entre conceptos (por ejemplo "más genérico que" o "más específico que"). (UNE 50-106-90, 1990: 5)

Esta definición nos presenta el tesoro como una estructura organizada a partir de relaciones conceptuales explícitas, que incluye control de vocabulario, y cuya función es la indización.

Por otra parte, la norma elaborada por la NISO (National Information Standards Organization), considera que un tesoro es:

a controlled vocabulary of terms in natural language that are designed for postcoordination
(ANSI/NISO Z39.19-2003, 2003: 1)

Esta definición agrega la característica del uso postcoordinado de los tesauros; o sea, la exigencia de que sus términos sean interrelacionados en el momento de la búsqueda de información. Si bien en el sitio web de la NISO, encontramos esta otra definición de

⁴⁰⁸ WERSIG, G. *Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis*. München: Verlag Dokumentation Saur KG, 1978.

⁴⁰⁹ AITCHISON, J. ; BAWDEN, D. ; GILCHRIST, A. *Thesaurus Construction and Use: A Practical Manual*. Chicago/London: Fitzroy Dearborn, 1971.

tesauro que enfatiza la labor de control de vocabulario y la organización estructurada de dicha herramienta⁴¹⁰:

Una colección de palabras o frases con términos equivalentes explícitamente identificados y con palabras y frases, por ejemplo los homógrafos, identificados unívocamente. Esta colección de términos puede incluir relaciones jerárquicas o de otro tipo.

Si consideramos las definiciones propuestas por los manuales de mayor difusión en el ámbito de la representación y recuperación de información, cabe destacar la de Aitchinson y Gilchrist, quienes definen tesauro como un vocabulario de un lenguaje controlado de indización, organizado formalmente de manera que las relaciones conceptuales son establecidas a priori, y que puede ser utilizado para la recuperación de información. Con esta definición, sumamos explícitamente otra de sus funciones: la recuperación de información.

Por su parte, van Slype considera que el tesauro es⁴¹¹:

una lista estructurada de conceptos, destinados a representar de manera unívoca el contenido de los documentos y de las consultas dentro de un sistema documental determinado, y a ayudar al usuario en la indización de los documentos y de las consultas.

Este autor, si bien presenta características ya mencionadas: control de vocabulario, estructuración conceptual, uso en la indización y recuperación de información, amplía la definición al introducir al usuario como beneficiario de dicha herramienta.

A partir de las características que hemos resaltado de las definiciones anteriores, se puede establecer una primera conceptualización de lo que es un tesauro bajo nuestro punto de vista. Un tesauro es un tipo de lenguaje documental que representa la estructuración conceptual de un determinado campo del conocimiento.

⁴¹⁰ ANSI/NISO Z39.19-2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. En: <http://webs.um.es/isgil/Z39-19-2005.pdf>. Consultado el 15/12/2009.

⁴¹¹ SLYPE, G.van. *Lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Peñaranda : Fundación Germán Sánchez Ruipérez, 1991. p.115.

Por tanto, desde su concepción etimológica, los tesauros no son más que diccionarios en los cuales cada entrada es catalogada junto con sus sinónimos⁴¹² como si se tratara de una red semántica o de una recuperación de términos por cluster. Esto lo vemos claramente en el siguiente ejemplo de recuperación en una herramienta denominada Visual Thesaurus. En este caso hemos buscado el descriptor “Subject”, como vemos se recupera éste en relación con otros descriptores, como topic, constituyendo un cluster de “Subject”. Cluster que se corresponderá, como veremos más adelante, con el Synset de la red semántica Wordnet.

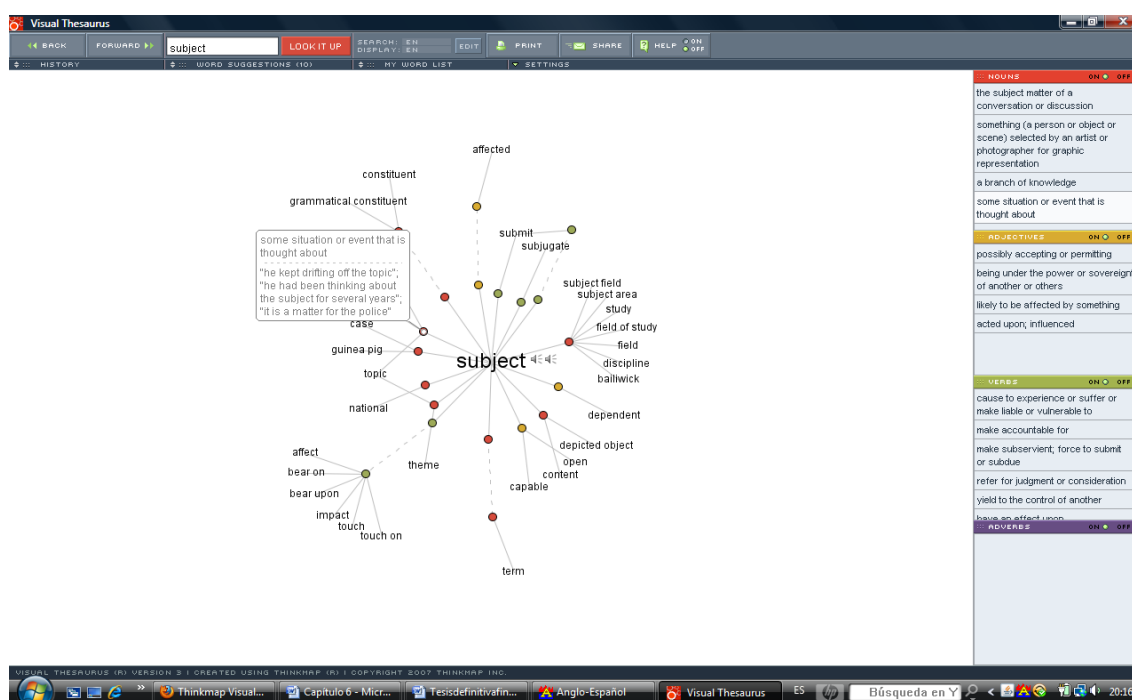


Figura 25. Buscador Visual Thesaurus.

Los primeros sistemas automáticos de indización comenzaron haciendo uso de uniterminos. Esto permitió a Taube diseñar el sistema Uniterm, compuesto por una colección de registros, donde cada uno de ellos contenía una única denominación y los números de aquellos documentos asociados de forma unívoca con cada una de las palabras. Uniterm se basaba en la hipótesis de que cada idea podría ser representada por una única palabra.

Su empleo surgió como respuesta a la necesidad de crear una lista de términos autorizados para la indización de literatura científica, ya que se trata de un lenguaje combinatorio constituido por listas de términos que representan un ámbito científico así como una serie

⁴¹² CURRÁS, E. *Ontologías, taxonomías y tesauros. Manual de construcción y uso*. Gijón : Trea, 2005. p.78.

de relaciones semánticas que le proporcionan gran capacidad de recuperación documental⁴¹³. Comenzó a ser utilizado para designar un diccionario que recogía todas las palabras y frases de la lengua inglesa, pero con la novedad que se trataba de un vocabulario organizado por significados y no ordenado por orden alfabético. Así, se estableció un precedente para vocabularios que relacionan sus términos entre sí a través de clases de relaciones semánticas predefinidas.

El programa Unisist⁴¹⁴ definió el término tesoro para documentación bajo dos parámetros: el estructural y el funcional. En el primer caso entendería el tesoro como un vocabulario controlado dinámico de términos relacionados semánticamente entre sí y por relaciones genéricas que cubren un dominio de conocimiento específico. Bajo el segundo parámetro, vendría a ser una herramienta de control terminológico empleado en la traducción del lenguaje natural de los documentos, de los indizadores o de los usuarios, a un lenguaje más restringido.

Desde las primeras iniciativas como el Roget's hasta la actualidad, el tesoro ha ido desarrollando su marco definitorio y su construcción teórica y metodológica, a través de la introducción de nuevos modelos cognitivos y aproximaciones centradas en el usuario. Así pues, resultado de esta evolución⁴¹⁵, es la concepción del mismo como un lenguaje postcoordinado, normalizado y especializado empleado con fines documentales, donde los elementos lingüísticos que lo componen, términos simples o compuestos, se encuentran relacionados entre sí sintáctica y semánticamente.

Lenguaje especializado por ser empleado en un dominio restringido. Por normalizado se entiende un lenguaje controlado donde las unidades lingüísticas son términos. Por lenguaje postcoordinado se entiende que los términos se combinan en el momento de su empleo en oposición a los lenguajes de indización precoordinados donde los términos designan materias complejas que son coordinadas previamente a su empleo.

La definición normalizada acerca de los tesauros fue establecida por la NISO⁴¹⁶ en un documento que establece las normas para la construcción, formato y gestión de tesauros⁴¹⁷,

⁴¹³ GIL URDICIÁIN, B. *Manual de lenguajes documentales*. Gijón : Trea, 2004, p.183.

⁴¹⁴ UNESCO. *Guidelines for the establishment and development of monolingual thesauri*. Paris, 1973.

⁴¹⁵ CURRÁS, E. *Ontologías, taxonomía y tesauros. Manual de construcción y uso*. Gijón: Trea, 2005.

⁴¹⁶ NISO (National Information Standards Organization)

donde se define el tesauro como un vocabulario controlado organizado en un orden definido y estructurado para que las relaciones de equivalencia, homógrafas, jerárquicas y asociativas entre los términos sean dispuestas de forma clara e identificadas por indicadores normalizados de relaciones que son empleados recíprocamente.

La norma internacional 2788-1986, publicada por AENOR en su correspondiente versión en lengua española⁴¹⁸, define el tesauro de acuerdo con dos criterios⁴¹⁹, su función:

El tesauro es un instrumento de control terminológico que traduce a un lenguaje sitémico o documental el lenguaje natural empleado en los documentos y por los usuarios.

y su estructura:

el tesauro consiste en un vocabulario controlado y dinámico de términos relacionados semántica y jerárquicamente, que se aplica a un campo específico del conocimiento.

Los tesauros especializados están organizados por medio de relaciones que forman una estructura jerárquica entre sus términos. El tipo de relación más frecuente en el tesauro es la jerárquica de genérico/específico, mientras que las relaciones no jerárquicas incluyen relaciones de equivalencia y asociación. Los tesauros monolingües son estructuras muy jerárquicas en las que un término sólo tiene un término genérico, pero en los tesauros multilingües a menudo se permite que existan varios genéricos.

Los tesauros son usados ampliamente para la indización y la catalogación en centros de documentación, especialmente cuando se está tratando con dominios restringidos donde se necesita de un elevado grado de detalle. En el ámbito jurídico destaca el empleo que se hace del tesauro EUROVOC, editado por la Unión Europea para homogeneizar el vocabulario jurídico a fin de facilitar la traducción de sus directivas y trasposición a cada uno de los estados miembro. Este tesauro es tomado como punto de partida por muchos organismos del estado español para la constitución de sus propios tesauros, así por ejemplo tanto el Senado como el CEPC tienen versiones adaptadas del mismo.

⁴¹⁷ ANSI/NISO Z39.19:1993 American national Standard guidelines for thesaurus structure, construction and use. New York, 1993.

⁴¹⁸ ISO. Documentación. Directrices para el establecimiento y desarrollo de tesauros monolingües. Norma internacional ISO 2788-1986. *Revista Española de Documentación Científica*, vol.12, nº 4, 1989. p.464.

⁴¹⁹ GIL URDICIÁN, B. Op. cit. p.183.

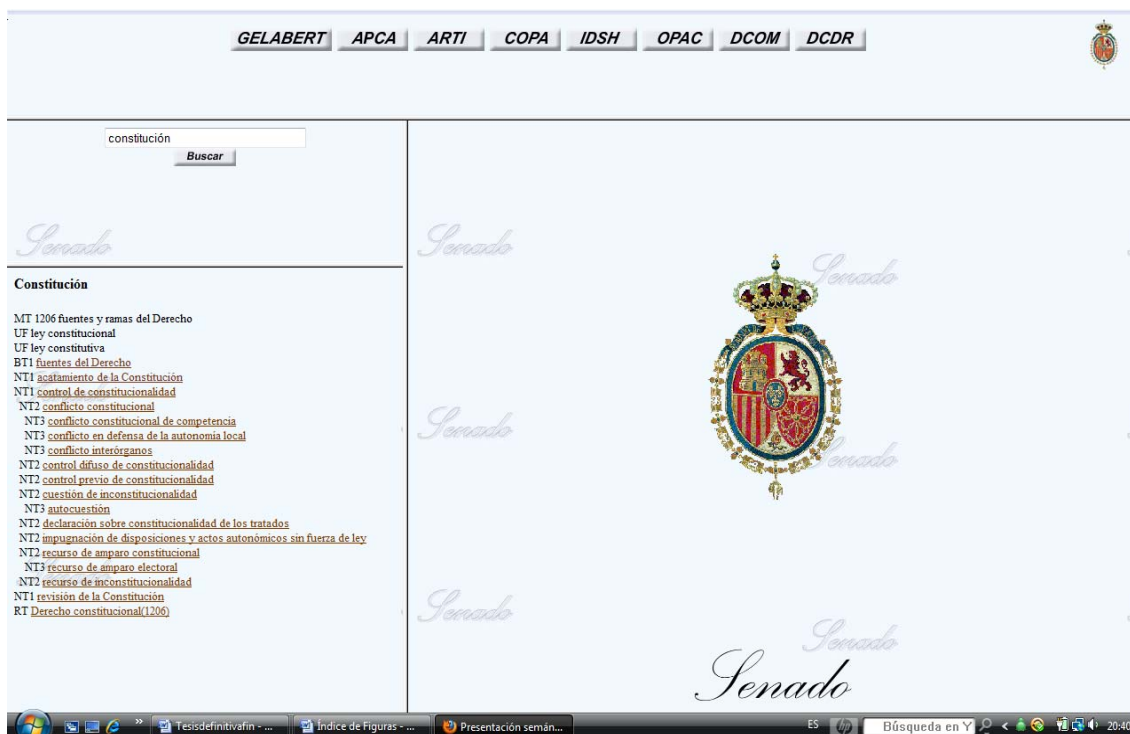


Figura 26. Tesauro del Senado. Ejemplo de adaptación del EUROVOC.

En cuanto a las funciones del tesauro, pasan por entender que éste⁴²⁰ fue adoptado por la disciplina de la Documentación por la forma de organización del vocabulario de indización y recuperación. Un tesauro puede funcionar bien en un entorno organizativo para la representación de materias así como en búsquedas informativas. La representación de las materias de los documentos la efectúa el indizador, quien analiza el documento, identifica sus contenidos y los traduce en términos permitidos por el tesauro.

Un tesauro puede ser empleado no sólo para ayudar en la elaboración de consultas efectuadas por el usuario en las búsquedas sino además por el indizador durante el proceso de clasificación. Tanto para el indizador como para el usuario, el tesauro por medio de su estructura de términos y relaciones, proporciona una asistencia para localizar los términos más adecuados a una materia. Por tanto, el tesauro es un componente fundamental en los sistemas de recuperación en tanto que efectúa las tareas de: a) determinación del término que puede ser empleado en el proceso de indización, b) determinación del término más adecuado a la búsqueda para obtener un resultado pertinente, y c) permite la introducción de nuevos términos en su estructura con el objeto de acercar el lenguaje del usuario al lenguaje del sistema y adaptar los sentidos semánticos a los propósitos de recuperación.

⁴²⁰ CURRÁS, E. *Tesauros: manual de construcción y uso*. Madrid : Kaher II, 1998.

Lancaster destaca la dificultad de mostrar con precisión la evolución del tesoro⁴²¹, una vez que la sucesión de influencias no está clara. La evolución histórica del tesoro se divide en dos líneas: la basada en el desarrollo del sistema Uniterm creado por Taube en 1951, y la aparecida a partir del desarrollo de la Teoría de Clasificación Facetada. La primera de las dos líneas deriva de la aproximación alfabética estadounidense y la segunda línea deriva de la clasificación bibliográfica europea y específicamente del Reino Unido, influida por la teoría de Ranganathan. La investigación analítico-sintética de Ranganathan establecía bases teóricas sólidas para los métodos de clasificación, con implicaciones sobre la indización alfabética de materias y la construcción de tesauros, de los tesauros facetados. Ambas líneas de evolución de los tesauros, que tienden a converger en la norma ISO 2788, en la norma británica BS 5723 y en la norteamericana NISO Z39-19, poseen algunas diferencias destacadas debido a las diferentes motivaciones para su desarrollo. La línea americana adopta una línea evolutiva más pragmática motivada por la necesidad de mejora de las limitaciones del lenguaje documental Uniterm.

Por otra parte, la corriente europea del tesoro facetado, al estar fundamentado en la teoría clasificatoria de Ranganathan⁴²², aplica el uso de categorías para la organización de los conceptos en un dominio. El empleo de categorías para agrupar los conceptos permite una mejor organización de las jerarquías y un mejor posicionamiento de los términos asociados con los conceptos.

Otra corriente de evolución más reciente es la referida al tesoro basado en conceptos, también conocida como corriente de los tesauros terminológicos. Este tipo de tesauros⁴²³ surge de la convergencia de la Teoría Conceptual desarrollada por Dahlberg con la Teoría de la Clasificación. La principal contribución de la Teoría Conceptual es una mejor comprensión de la relación entre concepto y término, la organización de los conceptos a través de categorías y el empleo de definiciones para la correcta inserción de un concepto en el tesoro.

Los tesauros han sido también empleados en recuperación de información desde que se formuló el campo de “thesaurus approach” a las bases de datos científicas y jurídicas. La

⁴²¹ LANCASTER, F. W. *Indexing and Abstracting in Theory and Practice*. London : Facet Publishing , 2003.

⁴²² LANCASTER, F.W. *El control del vocabulario en la recuperación de información*. Valencia: Universitat de Valencia, 2002.

⁴²³ CAMPOS, M.L.A. *Linguagem documentária: teorias que fundamentam sua elaboracao*. Niteroi: Eduff, 2001.

idea era combinar el método de la clasificación con el método de la indización mediante la agrupación de palabras en familias conceptuales. Estas familias de agrupaciones de palabras-clave son tenidas en cuenta para definir una clasificación. Así mediante el empleo de un tesoro con sus correspondientes relaciones jerárquicas uno podría efectuar una búsqueda inclusiva entendida como la posibilidad de propagar automáticamente la asignación de palabras-clave por medio de la estructura jerárquica del tesoro. Los resultados esperados son así más fáciles de indizar para el indizador⁴²⁴ y facilita la recuperación para el usuario⁴²⁵.

En los últimos años, los tesauros y las estructuras tesaurales han ido ocupando un papel central en la recuperación de información digital, especialmente a raíz de la evolución de la red hacia la Web Semántica, dado el necesario uso que tiene de los lenguajes controlados para la recuperación por semántica y no por sintaxis. Pensemos que el fundamento de la recuperación de información digital se encuentra en los índices, pero estos están constituidos por palabras que no siempre tienen un significado propio. Sin embargo, el concepto de índice es muy genérico y permite incluir en su concepción vocabularios más complejos. Así por ejemplo, un glosario es básicamente una lista de términos y definiciones pensadas como un tipo de índice en el cual sólo se recoge un único tipo de occurrence, aquella que proporciona una definición que además se muestra en la misma página que el topic sin necesidad de tener que recurrir a un localizador.

Un tesoro, hace énfasis en un conjunto de características añadidas al índice. Fundamentalmente es una red de términos interrelacionados en un dominio particular, y aunque a menudo contiene información aclaratoria como definiciones o ejemplos de uso, la característica clave de un tesoro son las relaciones (associations) entre términos. Es decir, si escogemos un término, el tesoro nos indicará que otros términos son sinónimos, cuáles denotan una categoría genérica del mismo tipo, bien una categoría específica o bien una categoría relacionada. La característica con más potencial en la recuperación de los tesauros es que las relaciones (associations) están tipificadas, son agrupables en categorías de relaciones: jerárquicas, de equivalencia y asociativas. Esto es muy importante dado que hace posible no sólo decir que dos términos están relacionados, sino también “cómo” o “por

⁴²⁴ Hay más términos de indización para escoger.

⁴²⁵ Puesto que los términos de indización son agrupados y organizados en una estructura y por tanto hay más posibilidades para realizar una búsqueda exhaustiva.

qué” lo están, y así agruparlos, facilitando notablemente la navegación por el tesoro. Los tipos de asociaciones que se pueden emplear en un tesoro se encuentran normalizadas en las normas ISO 2788-1986 y su equivalente británica BS 5723:1987 para los tesoros monolingües, y la norma ISO5964-1985 y su equivalente BS 6723:1985 para los tesoros multilingües⁴²⁶.

El tesoro proporciona una organización semántica principalmente a través de la explicitación de las relaciones establecidas entre dichos conceptos y, eventualmente, a través de un significado restringido de los términos que los representan. Los tesoros, para la Documentación, son por lo tanto una herramienta de control terminológico, puesto que la estructuración del tesoro es en base a conceptos. Los conceptos, son representados por términos seleccionados que evidencian ese control terminológico. Con dicho control se busca neutralizar la sinonimia y la polisemia, ambas características naturales de la lengua, que dificultan la precisión de la indización y recuperación de la información, ambas funciones básicas de los tesoros. Los tesoros son herramientas construidas para el auxilio tanto de los profesionales de la información como de los usuarios finales.

El impacto tecnológico del entorno digital ha afectado por igual a recursos, herramientas y comportamiento del usuario en relación a la representación y recuperación de información. Esta influencia abre nuevas posibilidades en cuanto a las pautas de diseño, elaboración, gestión y uso de las herramientas utilizadas para la representación y recuperación de información. Desde esta perspectiva López-Huertas ha empezado a perfilar los aportes del entorno digital a los tesoros, que se pueden resumir en los siguientes elementos.

Un primer elemento a tener en cuenta es el enriquecimiento de la funcionalidad de la estructura de los tesoros a partir de la hipertextualidad. Esto se traduce en el establecimiento de hipervínculos, entre todos los elementos estructurales: descriptores, no descriptores, notas de alcance, así como entre las distintas partes del tesoro.

Un segundo elemento es la reducción de costos de actualización y mantenimiento. Debido a la creciente informatización de los procesos de construcción de tesoros y al progresivo

⁴²⁶ Actualmente acaba de aprobarse la norma británica BS8723 para la interoperabilidad de vocabularios estructurados que aglutina tanto los tesoros como las ontologías o SKOS.

abandono del soporte papel para la publicación de dichas herramientas, es viable una reducción de costos.

Un tercer elemento, es la integración del usuario al proceso de creación, gestión y optimización de los tesauros, a través de tests de usabilidad, uso de técnicas de modelado de usuario, etc. Esto permite elaborar herramientas que tengan en cuenta los requerimientos de los usuarios, y descartar su construcción como simples estructuras teóricas.

Un cuarto elemento es la posibilidad de aplicar medidas de reutilización e interoperabilidad en el momento de planificar y construir los tesauros. Con ello se posibilita el aprovechamiento y enriquecimiento de la información conceptual y lingüística ya generada para otros recursos.

Desde nuestro punto de vista, este último elemento es clave para la nueva generación de tesauros digitales. El uso de información conceptual y lingüística almacenada en otros tipos de recursos, como una ontología, permite el enriquecimiento a nivel estructural de los tesauros, y posibilita el incremento de la usabilidad de estas herramientas de cara a su utilización por parte de un usuario final no especializado.

4.3.3 Redes semánticas.

Una red semántica es una notación gráfica para la representación de conocimiento en modelos de nodos interconectados y arcos. En la práctica es un grafo constituido por nodos representando conceptos y enlaces representando relaciones entre los conceptos. Las redes semánticas se parecen a los tesauros en tanto que organizan un dominio explicitando las relaciones entre los términos del dominio, pero su estructura puede llegar a ser más compleja que la del tesoro. Además, las propiedades y relaciones de los nodos en la red son expresadas en lenguaje lógico. Una red semántica puede contener virtualmente cualquier tipo de relación pero normalmente los conceptos están organizados de acuerdo con los niveles de generalidad sobre la base de distinciones taxonómicas. Esa jerarquía es a menudo denominada de “tipo” o taxonómica o jerarquía del todo-parte.

Las redes semánticas han sido primeramente estudiadas y desarrolladas dentro del marco de la informática, la inteligencia artificial, la representación del conocimiento y la lógica, donde

el énfasis se ha puesto en la capacidad de razonamiento del sistema. El tipo de razonamiento más comunmente aplicado ha sido el mecanismo de inferencia. Su desarrollo ha sido especialmente intenso desde los años 60 con los primeros sistemas de traducción automática, hasta los años 90 cuando perdió impulso la investigación en inteligencia artificial simbólica.

Una de las principales aportaciones de las redes semánticas ha sido la creación de la ontología lingüística Wordnet, organizada a modo de red semántica.

Los índices, glosarios y tesauros son formas de mapeo de estructuras cognitivas que existen de forma expresa en los documentos. En el campo de la inteligencia artificial también existe la necesidad de ser capaces de representar el conocimiento a fin de apoyar la comunicación entre personas y máquinas. Un formalismo muy empleado de representación del conocimiento es el referido a los grafos conceptuales, en donde los elementos de construcción son los conceptos y las relaciones conceptuales. Así por ejemplo si tenemos el siguiente grafo conceptual para la frase “la soberanía dimana del pueblo”, tanto “soberanía” como “dimanar” y “pueblo” son conceptos, en tanto que se establece una relación asociativa entre un concepto con función de elemento agente y otro elemento con función de objeto. Esta relación podría ser establecida gráficamente de la siguiente manera: soberanía <- (agente) <- dimanar -> (objeto) -> pueblo. A las estructuras gráficas de este tipo se las conoce de distintas maneras dentro del campo de la inteligencia artificial como “redes semánticas”, “redes asociativas” o “mapas conceptuales”.

Dado que el modelo básico de redes semánticas es muy similar al referido a los topics y associations encontrados en los índices, si se combinan ambas aproximaciones sería posible crear un modelo muy eficaz en la gestión de la información y del conocimiento. Este modelo es lo que constituye la norma de los *topic maps*.

Si se añade el factor de las occurrences al modelo topic/association, los topic maps proporcionan un modelo para superar el hueco semántico como el que existe entre la representación del conocimiento y la gestión de la información.

La gestión del conocimiento es hoy en día un concepto muy empleado tanto en el ámbito empresarial como en el académico debido a la necesidad de facilitar el acceso ordenado a

los recursos de información que Internet pone a disposición del usuario. Para las grandes compañías de consultoría, la gestión del conocimiento trata fundamentalmente acerca de nuevas técnicas de gestión empresarial diseñadas sobre la base de que las personas constituyen el activo primario en una economía basada cada vez más en el conocimiento. Para otras compañías la gestión del conocimiento es equiparable a la gestión de información. Sin embargo, el conocimiento es muy diferente de la información. La diferencia está en que entre conocer una cosa es distinto que tener información sobre la misma. Según Ruggles, la gestión del conocimiento supone tres tareas esenciales: generación, codificación y transferencia⁴²⁷, por tanto los topic maps pueden ser considerados como una norma para la codificación, requisito necesario para el desarrollo de herramientas de asistencia en la generación y transferencia de conocimiento.

4.3.4 Ontologías.

La ontología⁴²⁸ es una disciplina de la Filosofía que trata sobre el ser y la naturaleza de su existencia, qué tipos de cosas existen y qué entidades hay en el Universo. La palabra ontología significa para la Real Academia Española⁴²⁹:

Parte de la metafísica que trata del ser en general y de sus propiedades trascendentales.

Sin embargo, el aspecto que interesa tratar de las Ontologías aquí es el recogido por Sowa en su obra ya clásica en todos los trabajos referidos a la Web Semántica. Para Sowa, una Ontología es⁴³⁰:

el estudio de las categorías de cosas que existen o pueden existir en algún dominio. El producto de tal estudio, denominado una Ontología, es un catálogo de los tipos de cosas que se asume que existen en un dominio de interés D desde la perspectiva de una persona que emplea un lenguaje L con el propósito de hablar acerca del dominio D.

Por lo tanto, en Documentación, una Ontología es una especificación explícita de una conceptualización que incluye un vocabulario de términos⁴³¹; es decir, el modelo de trabajo

⁴²⁷ RUGGLES, Rudy L., (ed). *Knowledge management tools*. Boston : Butterworth-Heinemann, 1997. pp.23-24.

⁴²⁸ GÓMEZ-PÉREZ, A. ; FERNÁNDEZ LÓPEZ, M. ; CORCHO, O. *Ontological Engineering*. London : Springer, 2004, p.6.

⁴²⁹ RAE. En: www.rae.es. Consultado el 15/07/2009.

⁴³⁰ SOWA, John F. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. New York : Brooks Cole, 2000. p.492

de entidades e interacciones en un dominio particular de conocimiento, por ejemplo el caso de una Ontología construida para la realización de una Base de Datos referida a Biblioteconomía y Documentación como puede ser el caso de LISA. En esta base de datos hay un vocabulario específico del área como pueda ser: indización, clasificación, recuperación; hay unas relaciones entre los términos muy específicas y sobre todo existen unas fuentes sobre la materia a recuperar. Todo esto en conjunto representa un modelo o esquema conceptual que pretende reflejar una realidad de un dominio de conocimiento muy particular.

No existe acuerdo entre los autores acerca de la fecha en que se comenzó a hablar de ontologías en el ámbito de las tecnologías de la información, pero parece que hacia los años 80 se admite la conceptualización de un dominio como parte necesaria para la adquisición de conocimiento. En el ámbito de los sistemas expertos, es donde se empieza a admitir que la suma de conocimientos en un sistema experto, cuando se codifican, pueden llegar a comportarse como un sistema humano⁴³². A partir de 1993, Gruber, pionero en el empleo de la web para el intercambio de conocimiento, definió una ontología como una descripción de los conceptos y relaciones que pueden existir para un agente o comunidad de agentes⁴³³.

La definición canónica aceptada de forma más generalizada entre los autores es la proporcionada por el consorcio W3C y recogida por King y Reinold⁴³⁴:

Una ontología define los términos empleados para describir y representar un área de conocimiento. La Web Semántica necesita ontologías con un grado significativo de estructuración. Estas deben especificar descripciones para los siguientes tipos de conceptos: i) Clases en sus distintos dominios de interés; ii) relaciones que puedan existir entre las anteriores; iii) las propiedades de las clases.

¿Pero qué diferencias o valores añadidos aporta respecto de los lenguajes documentales que hemos venido utilizando, como los tesauros o las taxonomías? Pues que ni las redes

⁴³¹ MOREIRO GONZÁLEZ, J.A. *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural*. Gijón : Trea, 2004, p.215.

⁴³² CURRÁS, E. *Ontologías, taxonomía y tesauros. Manual de construcción y uso*. Gijón : Trea, 2005, p.35.

⁴³³ GRUBER T.R. A translation approach to portable ontologies. *Knowledge Acquisition*, vol.5, nº 2, 1993. p.200.

⁴³⁴ KING, B. ; REINOLD, K. *Finding the concept, not just the word: a librarian's guide to ontologies and semantics*. Oxford : Chandos Publishing, 2008. p.8.

semánticas, ni los tesauros, ni las taxonomías ofrecen definiciones, relaciones ni clasificaciones al mismo tiempo⁴³⁵. Las ontologías combinan los elementos de estos tres tipos de lenguajes documentales, proporcionando conceptos como una red semántica, estableciendo relaciones como un tesoro y proporcionando categorizaciones como una taxonomía. De esta manera permite que un concepto o *topic* pueda ser explorado desde todos los ángulos por medio del examen automático de todos los tipos de relaciones que el cerebro humano realiza. Es más, la ontología permite a su autor la definición de un número ilimitado de relaciones, entendidas éstas como conceptos a ser representados por frases en vez de únicamente palabras. Esto permite generar un sitio web con infinidad de relaciones que no pueden ser expresadas de manera explícita con palabras y términos; es decir, estamos representando contenidos implícitos en la relación entre conceptos.

Así por ejemplo, en el *topic map* que hemos construido para el sitio web del CEPC, no se dice de manera explícita que la Directora General se encuentre en el organigrama del centro, pero en el mapa conceptual generado automáticamente a partir del topic map se visualiza claramente que el *topic* “Directora General” es una instancia del concepto Organigrama, pero es que además sabemos que topic organigrama tiene un topic role que es organizar, lo cual no se explicita en ninguno de los topic anteriores.

Por otro lado, las ontologías son capaces de reconocer las distintas maneras con las que se puede recoger una misma idea, pero siempre en el ámbito de un dominio de conocimiento y con el lenguaje de especialidad empleado para la designación de los topics y sus relaciones. Esta capacidad para adaptar los contenidos a los perfiles de los usuarios es una propiedad que permite recoger las distintas opciones de acceso a la información para la diversidad de sus usuarios⁴³⁶, algo que no permiten los lenguajes documentales, a excepción de los tesauros facetados.

⁴³⁵ Ibid., p.10.

⁴³⁶ MARCOS, M.C. *Interacción en interfaces de recuperación de información: conceptos, metáforas y visualización*. Gijón : TREA, 2004. p.282.

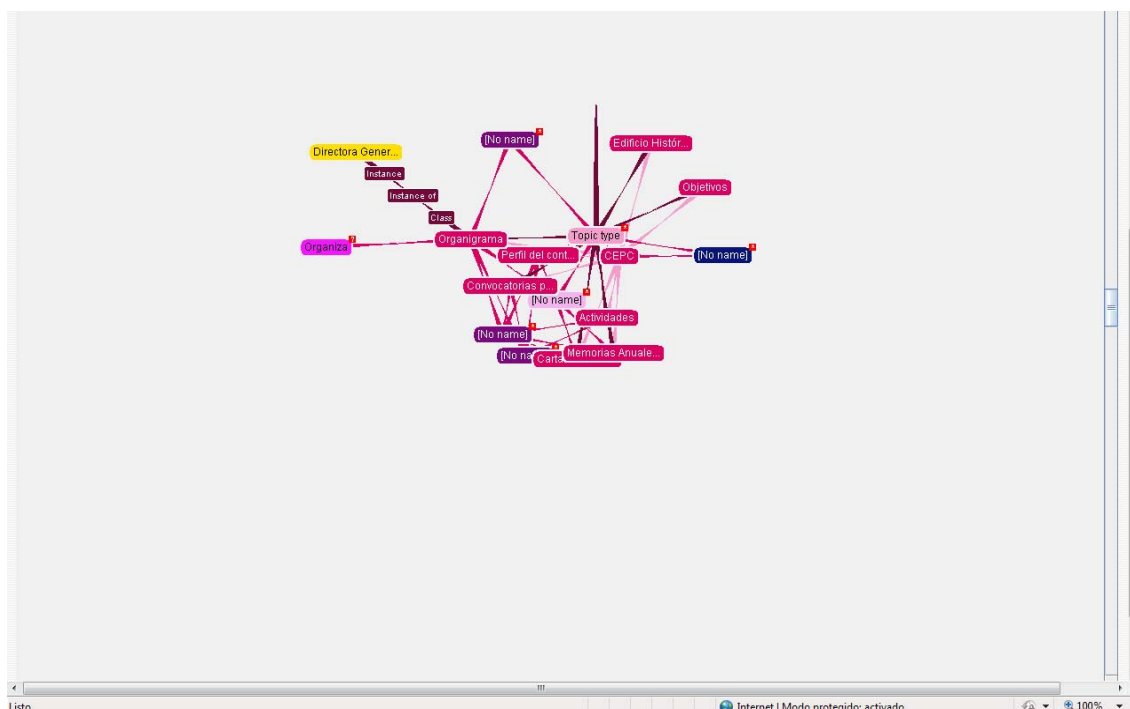


Figura 27. Ejemplo de representación de relaciones implícitas entre conceptos en el topic map del CEPC.

4.4 Topic maps y Esquemas de Organización Semántica de la Información.

4.4.1 Introducción.

Los topic maps tratan del modelado de las relaciones en el mundo real entre objetos, personas y conceptos⁴³⁷. El ser humano está dotado de una capacidad de conocimiento única de la situación y el contexto que gobierna cómo interpretamos e interactuamos con nuestro entorno. Esto se aplica tanto al espacio geofísico como al espacio social y en un contexto particular al espacio de información. A partir de esta base conceptual podemos definir el propósito primario de un topic map como el modo de proporcionar un entorno contextualmente rico donde el espectro de materias de una colección de recursos informativos puede hacerse interpretable y navegable. Así, para los profesionales de la información un topic map es fundamentalmente una herramienta de navegación⁴³⁸.

⁴³⁷ BIEZUNSKI, M. Introduction to the Topic Maps Paradigm. En PARK, J. *XML Topic Maps. Creating and using Topic Maps for the Web*. Boston : Addison-Wesley, 2002. pp.18ss.

⁴³⁸ BATER, B. Topic Maps: Indexing in 3-D. En GILCHRIST, A. ; MAHON, B. *Information architecture: designing information environments for purpose*. New York : Neal-Schuman Publishers, 2004. p.134.

Es por esto último por lo que debemos situar los topic maps en el campo de la arquitectura de información, campo interdisciplinar cuyo objetivo es definir las posibilidades de organización del conocimiento en relación con los recursos de información de un dominio particular para su gestión⁴³⁹. En la mayoría de las organizaciones existe una estructura que emplea simultáneamente diversos recursos informativos: documentos digitales en diferentes formatos, bases de datos, canales de noticias y documentos en papel. El acceso sincrónico a todos estos documentos se realiza a través de navegadores y motores de búsqueda para cada sistema propietario, con la complejidad y gastos que ello supone. Por ello, la tendencia general ha sido migrar todos los documentos a un formato web estándar, XML. Sin embargo, los motores de búsqueda han sido diseñados para la recuperación exhaustiva y homogénea de recursos en la web, haciendo que éstos sean muy dependientes de los sitios web. Así, cuando se contaba con archivos de texto que podían ser indizados, los motores de búsqueda no eran capaces de obtener información acerca de los registros almacenados en una base de datos o en formatos no propietarios de texto. Esto hacía que la información fuera poco interoperable tanto entre sí como con otras arquitecturas de información, y desde luego dificultaba la operatividad de cualquier sistema de gestión del conocimiento fuera de los formatos para los que son creados.

Una de las primeras soluciones a este problema fue la implantación de una sintaxis común, pero una segunda solución vino de los portales en entornos corporativos, donde la inserción de motores de búsqueda con la opción de búsqueda avanzada, permitió mejorar considerablemente la recuperación de información con independencia de su formato. E incluso motores de búsqueda como Google permiten la inserción de comandos para la recuperación por un formato particular⁴⁴⁰.

El problema viene dado por la dificultad de recuperar documentos de los que se desconocen datos básicos como el autor, título o alguna palabra clave por la que haya podido ser descrito. Es la dificultad de la recuperación por la semántica, por el significado de sus contenidos. Los topic maps son una tecnología pensada para reconducir el problema de la caracterización semántica y la categorización de documentos y secciones de documentos Web respecto a su contenido intrínseco, frente a la tradicional idea de unos

⁴³⁹ MORVILLE, P. ; ROSENFELD, L. *Information Architecture*. Beijing : O'Reilly, 2006. p.4.

⁴⁴⁰ En Google con el comando "filetype" podemos exigir que nos recupere los documentos en formatos tales como pdf, xls, doc, o incluso imágenes en cualquiera de sus formatos (jpg, gif, bmp) y vídeos.

índices que describen los contenidos a través de una colección de descriptores enlazados, pero sin que exista una organización estructurada de los descriptores en relación con las materias. Sin que haya referencias cruzadas a otras materias, lo que permitiría enlazar no descriptores individuales sino contenidos más amplios entre sí, así como conceder la posibilidad de que los usuarios puedan realizar las consultas de manera interactiva combinando las materias de su interés; es decir que el usuario disponga de contextos con los que poder adaptar los índices a sus necesidades reales de consulta por contenidos⁴⁴¹. En este sentido podemos afirmar que los topic maps proporcionan acceso a la información existente en diferentes redes semánticas⁴⁴², lo que nos lleva a conceptualizarlos como metaíndices, o más bien un índice de índices. Son, por tanto, una tecnología web capaz de ofrecer un modo menos complejo y más interoperable de acceso a una colección heterogénea de recursos con independencia de su formato, ordenación y localización.

Aunque la aparición de los topic maps es novedosa, no lo es su conceptualización dado que es esencialmente un índice de los contenidos de una colección de recursos digitales de información. Precisamente al tratarse de un índice digital presenta numerosas ventajas *per se* respecto a un índice en papel a causa de la divergente dimensión del contexto informativo en formato digital respecto al papel. Los contenidos en papel a menudo son inadecuados para expresar sutilmente conocimientos expresados a través de la relación entre contenidos localizados en distintas fuentes de información, o sencillamente que no tienen una relación directa en el seno del mismo texto. Landow postula que⁴⁴³:

deben abandonarse los actuales sistemas conceptuales basados en nociones como centro, margen, jerarquía y linealidad y sustituirlos por otras de multilateralidad, nodos, nexos y redes

La información en papel nos atrapa en un entorno bidimensional y jerarquizado que dificulta la capacidad asociativa del pensamiento humano, en tanto que la información digital es un entorno informativo multidimensional. Esto nos permite sustraernos a la ordenación lineal de la narración tradicional para posibilitar una aprehensión de los

⁴⁴¹ DACONTA, M. ; OBRST, L. ; SMITH, K. *The Semantic Web. A guide to the future of XML, Web Services, and Knowledge Management*. Indianapolis : Wiley Publishing, 2003. p.167.

⁴⁴² MOREIRO GONZÁLEZ, J. A. *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural*. Gijón : TREA, 2004. p. 178.

⁴⁴³ LANDOW, G. *Teoría del hipertexto*. Barcelona : Paidós, 1995. p.14.

significados por medio de una red semántica de asociaciones que ayudan a delimitar el contexto interpretativo de la información.

No obstante la lectura multidimensional presenta numerosos problemas para el usuario de localización de la información que consulta en relación con otros documentos. El usuario se pierde con la sobrecarga de información, dificultando la navegación precisa y la asimilación contextual de los mensajes.

La necesidad de integrar el contexto en Documentación no es un problema nuevo. A comienzos de los años 70, se creó el proyecto PRECIS⁴⁴⁴ para el diseño de un sistema de indización que haciendo uso de la capacidad de procesamiento de los nuevos superordenadores, permitía generar unos índices permutados que preservaban el contexto semántico y sintáctico de las palabras clave⁴⁴⁵. Fue empleado por primera vez en la British National Bibliography en 1971 pero a pesar de haber suplantado un sistema más simple, COMPASS, fue retirado en 1996 ya que se consideraba demasiado complejo para el trabajo a mano; si bien, se debe decir que la complejidad no residía tanto en PRECIS sino en el hecho de aplicar la capacidad de los ordenadores a un espacio de información inherentemente limitado como es la bidimensionalidad de la información en papel.

El problema del contexto en el espacio bidimensional reside en la dificultad para realizar una conservación efectiva del mismo, problema que venía siendo reconocido desde hacía tiempo por los documentalistas. De hecho, ya entonces se comentaba que los sistemas de recuperación de información debían ser capaces de proporcionar más posibilidades para realizar búsquedas contextuales⁴⁴⁶. Incluso se llegó a abogar por sistemas diseñados para ser capaces de emplear una red de asociaciones entre elementos de información como medio para salvar sus diferencias semánticas⁴⁴⁷. En realidad, lo que se estaba diciendo era que el problema de la recuperación de información contextualizada podría ser reconducido cambiando la pregunta sobre: ¿qué es una necesidad de información?, a ¿por qué existe la necesidad de información?

⁴⁴⁴ AUSTIN, D. *PRECIS: a manual of concept analysis and subject indexing*. London : Council of the British National Bibliography, 1974.

⁴⁴⁵ BATER, B. Op. cit., p.144

⁴⁴⁶ ELLIS, D. A behavioural approach to information retrieval design. *Journal of Documentation*, vol.45, n° 3, 1989. p.193.

⁴⁴⁷ BELKIN, N. J., MARCHETTI, P. BRAQUE: design of an interface to support user interaction in information retrieval. *Information Processing and Management*, vol. 29, n°3, 1993, p.325.

Creemos que con los topic maps es posible responder a esta última pregunta ya que trata acerca del modelado de las relaciones entre conceptos y objetos desde una óptica dirigida por el usuario mediante el empleo de las facetas. Es decir, como veremos, los *topic maps* permiten a los usuarios por un lado situarlos en el contexto correspondiente a aquel en el cual se han generado sus demandas informativas, y por otro interrelacionar lenguajes documentales para obtener un lenguaje controlado de *topics* o materias⁴⁴⁸.

Su principal aportación es la integración contextual que hace de los lenguajes documentales entre sí, es un índice de índices, y con los documentos, además de sus respectivos metadatos. Metadatos que nos proporcionan una descripción de las propiedades intrínsecas de los documentos y son asimilables a las *occurrences* de los *topic maps*⁴⁴⁹.

Esto significa, que en la práctica, se facilita la localización precisa y pertinente no sólo de documentos sino sobre todo de objetos de información dispersos en documentos distintos. Así, por ejemplo, una posible aplicación de los topic maps como metaíndices generados a partir de una colección estructurada de índices es la reutilización de software al igual que se ha hecho con los tesauros⁴⁵⁰, dada la sistematización que hace de las relaciones entre los topics.

Al tratarse de metaíndices existe un elevado nivel de interrelación entre sus descriptores, lo que permite su interoperabilidad entre distintos campos de conocimiento. De hecho, un *topic map* incrementa el rango de relaciones de un sistema de organización del conocimiento hasta el punto de permitirnos hablar de ontologías⁴⁵¹ como categorías de objetos, esquema conceptual, que existen o pueden existir en un dominio de conocimiento y cuya finalidad es facilitar la comunicación y el reparto de la información entre diferentes entidades.

⁴⁴⁸ MOREIRO GONZÁLEZ, J. A. Op. Cit., p.178.

⁴⁴⁹ PEPPER, S. Expressing Dublin Core in Topic Maps. En MAITCHER, L. ; GARSHOL, L. *Scaling Topic Maps. Third International Conference on Topic Maps Research and Applications, TMRA 2007*. Berlin : Springer, 2007, p.188ss.

⁴⁵⁰ LLORENS MORILLO, J. Definición de una metodología y una estructura de repositorio orientadas a la reutilización : el tesoro de software. Director: Antonio Amescua Seco. Leganés : Universidad Carlos III, 1996.

⁴⁵¹ SOWA, J.F. *Knowledge representation : Logical, Philosophical and Computational Foundations*. New York : Brooks Cole, 2000. p.521.

Otra de las ventajas de los topic maps tiene que ver con su capacidad para mejorar el potencial semántico aportado por los metadatos⁴⁵². Hasta ahora nos encontrábamos con información acerca de las características de los documentos tales como su estructura o la autoría de los mismos, muchas veces repetida en varios lenguajes controlados. Por ello, contar con los documentos se hacía imprescindible ya que los sistemas de clasificación convencionales son jerárquicos y por tanto no permiten expresar toda la información necesaria acerca de la interrelación asociativa entre los contenidos. Con los *topic maps* desaparece esta limitación, y por tanto, ya no hace falta contar con los documentos porque su ausencia es suplida por la estructura navegacional del portal.

Así por ejemplo, en el portal de la biblioteca digital del New Zealand Text Centre⁴⁵³ (NZETC) se utiliza un *topic map* para estructurar los recursos del sitio de forma que estos puedan ser navegables sin tener que contar con los documentos en sí mismos. En este caso hemos seleccionado en el *topic type* author el topic Miriam A. Aiken. A partir de la *association* “mentioned in” podemos relacionar nuestro *topic type* “author” con el *topic type* “Works”, de forma que podemos recuperar todas las obras a texto completo de esta autora, que también son *topics*, con independencia de donde estos se encuentren.

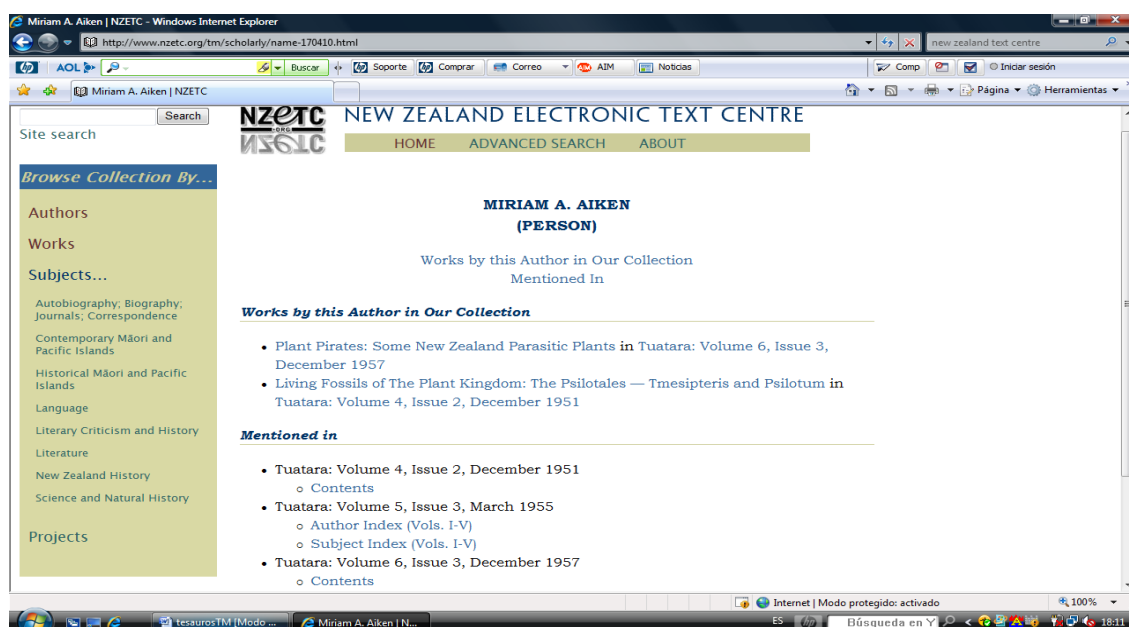


Figura 28. Ejemplo de búsqueda por autor empleando topic maps en el portal NZETC.

⁴⁵² GARSHOL, L.M. *Metadata? Thesauri? Taxonomies? Topic Maps. Making sense of it all.* En: <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html#ahmed> Consultado el 15/07/2009.

⁴⁵³ La biblioteca digital de acceso abierto del New Zealand Electronic Text Centre es una iniciativa de la Victoria University of Wellington para proporcionar textos y materiales de todo tipo acerca de la cultura de las islas del Pacífico. En: <http://www.nzetc.org/>. Consultado el 15/05/2009.

Como vemos, el empleo de *topic maps* para la organización de los recursos informativos de un sitio web, no sólo mejora las posibilidades de búsqueda por navegación con el empleo de una red semántica, sino que permite inferir más información de la explicitada en un corpus documental. Análogamente a lo que ocurre cuando se cambia el formato de un documento generado en Word a un formato en XML para permitir su difusión en más de un formato sin necesidad de modificar su estructura⁴⁵⁴.

Otro de los grandes beneficios aportados por el empleo de *topic maps*, es que se potencian las posibilidades para la localización oportuna de sitios web, a los que al mismo tiempo se les dota de fiabilidad, tanto por sus lenguajes de consulta específicos como por una visualización gráfica más usable y que rompe con los tradicionales esquemas lineales y jerárquicos de los lenguajes controlados⁴⁵⁵.

No obstante, los *topic maps* tienen algunos puntos débiles derivados de la inmadurez de sus principios, lo que supone algunas desventajas respecto a modelos de organización del conocimiento más consolidados como RDF. Algunos autores, Pepper o Garshol, comentan que los *topic maps* son un complemento de otros sistemas con los que debería integrarse, más que reemplazarlos, ya que resulta demasiado costoso traducir las colecciones de documentos en otros lenguajes al lenguaje de los *topic maps*⁴⁵⁶. Por ello sugieren que la estrategia más efectiva sería dar forma de topic map a la información en otros lenguajes para poder ser tratada con las técnicas de los *topic maps*⁴⁵⁷.

Otra desventaja de la aplicación de los topic maps proviene de los impulsores de esta iniciativa, dado que, muchos de ellos, provienen del ámbito de la computación y desconocen las posibilidades de las técnicas documentales a las que sin embargo aluden e intentan aplicar de manera persistente⁴⁵⁸. Incluso, en ocasiones hablan de los topic maps como un marco para la representación de características básicas de los lenguajes documentales a partir del cual se puede crear cualquier lenguaje documental. Y en efecto, es

⁴⁵⁴ GOLDFARB, CH. ; PRESCOD, P. *XML Handbook*. Upple Sadle River : Prentice Hall, 2002. p.32.

⁴⁵⁵ Ibid., p.639.

⁴⁵⁶ GARSHOL, L. *Living with topic maps and RDF*. En: http://www.ontopia.net/topic_maps/materials/tmrdf.html Consultado el 12/07/2009.

⁴⁵⁷ PASSIN, T. *Explorer's Guide to the Semantic Web*. Greenwich : Manning, 2004, p. 72.

⁴⁵⁸ Véase por ejemplo el artículo de Pepper, S. *The TAO of Topic Maps*. En: http://www.ontopia.net/topic_maps/materials/tao.html Consultado el 15/05/2009.

posible, pero los topic maps van más allá de un mero marco de representación de lenguajes documentales, que por otra parte ya ha sido creado y se denomina SKOS⁴⁵⁹.

Respecto a ese “más allá” de los topic maps podemos decir que permiten la generación de estructuras más ricas y que sean los propios usuarios quienes las definan por sí mismos. Esto significa que, por ejemplo, el esfuerzo necesario para construir un sistema de clasificación puede llegar a ser mucho más complejo, al tener que considerar un mayor rango de elementos, propiedades y relaciones. Hasta cierto punto, la existencia de ontologías predefinidas y reusables alivian esta situación, pero para aquellas personas no familiarizadas con las nuevas tecnologías y concretamente con aspectos tales como hacer que un ordenador personal se comporte como un servidor es un problema todavía real⁴⁶⁰.

4.4.2 Topic maps como modelo de organización de contenidos.

Desde hace años se viene trabajando con la posibilidad de aplicar lenguajes de marcado tales como SGML⁴⁶¹ y XML⁴⁶² para la organización y recuperación de documentos digitales. Sin embargo, estos lenguajes aunque facilitaban notablemente el manejo de grandes corpora textuales, especialmente si se trataba de documentación estructurada, no llegaban a resolver los problemas de búsqueda en corpora multimedia con técnicas de expansión de la consulta⁴⁶³.

Cuando se introducía XML en una organización se realizaba con dos objetivos: bien para estructurar su documentación o bien para hacer posible que las distintas aplicaciones de la organización se pudieran entender entre sí; es decir, XML resolvía los problemas de estructuración e interoperabilidad de la documentación con independencia de las aplicaciones con que esta hubiera sido creada⁴⁶⁴. No obstante, era incapaz de resolver los problemas de localización de la información porque lo que cambia con XML es que el procesamiento de documentos se vuelve más controlable, y con mayor grado de

⁴⁵⁹ SIMPLE KNOWLEDGE ORGANIZATION SYSTEM. En: www.w3.org/2004/02/skos/. Consultado el 15/05/2009.

⁴⁶⁰ Experiencia sufrida por el autor de esta tesis, quien lo aprendió en la Universidad de Brighton.

⁴⁶¹ GOLDFARB, Ch. *The SGML handbook*. Oxford : Clarendon Press, 1995.

⁴⁶² GOLDFARB, Ch. *Manual de XML*. Madrid : Prentice Hall, 1999.

⁴⁶³ TUDHOPE, D. ; ALANI, H. ; JONES, C. Augmenting thesaurus relationships: possibilities for retrieval. *The Journal of Digital Information*, vol.1, nº 8, p. 11. En: http://eprints.ecs.soton.ac.uk/4484/1/Tudhope_JoDI.pdf Consultado el 15/07/2009.

⁴⁶⁴ GOLDFARB, CH. ; PRESCOD, P. *XML handbook*. Upper Saddle River : Prentice Hall, 2002. pp.7ss.

automatización, además de facilitar la interoperabilidad entre las aplicaciones con las que se generan los documentos.

A partir de finales de los años 90 comienzan a difundirse una serie de tecnologías tanto entre el ámbito empresarial agrupado en consorcios como OASIS o el W3C, así como la constitución de grupos conjuntos de trabajo en torno a organismos internacionales de normalización tales como International Standard Organization para la adaptación de los lenguajes de marcado a las necesidades de expansión de la consulta y de recuperación semántica de la información. Entre estas iniciativas surge la propuesta de los topic maps en el seno del working group de lenguajes de marcado de ISO⁴⁶⁵.

Los topic maps son una norma ISO que tiene por objetivo la creación de un índice de información que se encuentra fuera del espacio informativo, a modo de mapa con el cual poder acceder por hipertexto a varias fuentes de información con independencia de donde éstas se encuentren⁴⁶⁶. Creándose así una división en dos niveles entre la organización de las fuentes de información que se encuentran en distintas bases de datos y servidores, correspondiente a los rectángulos de la parte inferior del diagrama, y el topic map, correspondiente a la nube de la parte superior del diagrama, que describe la información en los documentos y la organiza de forma coherente de acuerdo con un esquema conceptual. Pero estos dos niveles no se encuentran separados sino que se relacionan a través de hipertextos definidos por las URI's (líneas continuas en el diagrama), desde los descriptores o topics del nivel superior hasta los recursos informativos; con independencia del formato de éstos y de que se encuentren en una o muchas bases de datos, repositorios o servidores o en los tres al mismo tiempo⁴⁶⁷.

⁴⁶⁵ NEWCOMB, S. A perspective on the quest for global knowledge interchange. En PARK, J. *XML Topic Maps*. Boston : Addison-Wesley, 2003. p.37.

⁴⁶⁶ DACONTA, M. ; OBRST, L. ; SMITH, K. *The semantic web*. Indianapolis : Wiley Publishing, 2003. p.167.

⁴⁶⁷ GARSHOL, L. *What are Topic Maps*. En: http://www.xml.com/pub/a/2002/09/11/topic_maps.html Consultado el 15/07/2009.

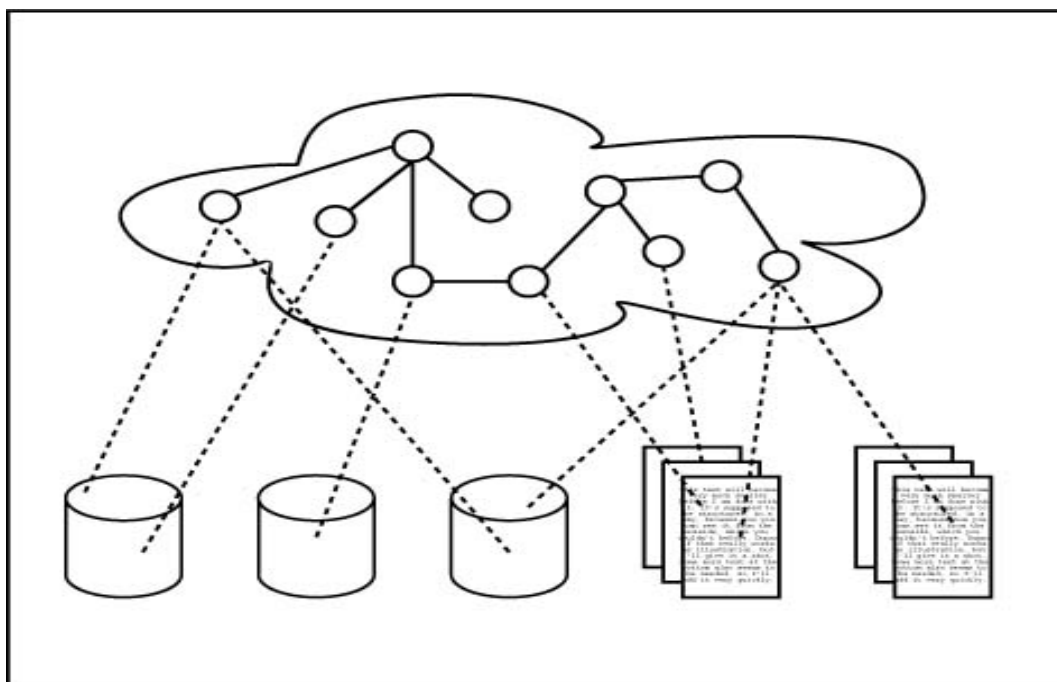


Figura 29. Representación gráfica de la relación de un topic map con los objetos de información según Garshol.

El *topic map* toma los conceptos clave descritos en las bases de datos y documentos y los relaciona entre sí con independencia de su situación física en las fuentes de información que están siendo indizadas. Así por ejemplo, cuando en un documento se dice que: “El proceso de mantenimiento para la parte X consta de los siguientes pasos...” el topic map permite decir: “La parte X es del tipo Q y está contenida en las partes Y y Z y su proceso de mantenimiento reside en el documento W”. Como vemos el topic map lo que hace es gestionar los contextos de la información más que gestionar únicamente la información⁴⁶⁸.

El resultado es una estructura de información que rompe con el tradicional corsé jerárquico hasta ahora empleado para deducir la semántica de los documentos. Un *topic map* normalmente contiene muchas jerarquías superpuestas e interrelacionadas por medio de enlaces cruzados que proporcionan múltiples significados, proporcionando así una visión poliédrica de los contenidos transmitidos por los documentos. Esto hace que la información sea mucho más fácil de localizar porque el usuario ya no actúa como los diseñadores de la información esperan que actúen sino que cuentan con múltiples estrategias de navegación que le conducen a los mismos resultados de búsqueda⁴⁶⁹.

⁴⁶⁸ NORDENG, T. ; GUESCINI, R. ; KARABEG, D. Topic maps for polyscopic structuring of information. *International Journal Context Engineering Education and Lifelong Learning*, vol.16, nº1/2, 2006. p.42.

⁴⁶⁹ BATER, B. Op. cit., p.139

El empleo más común que se ha hecho hasta ahora de los topic maps ha sido para la organización de contenidos en sitios web que faciliten la localización y comprensión de los contenidos demandados por el usuario. El topic map proporciona una estructura de sitio, y el contenido de sus páginas es interpretado en parte desde el topic map en sí mismo y en parte desde los documentos con los que se enlaza. Esta solución es idónea para portales, catálogos e índices de sitios, dado que se está representando el conocimiento contenido en los documentos al mismo tiempo que se proporciona una herramienta para la gestión del conocimiento. Además pueden ser empleados para la organización de contenidos en sistemas de gestión de contenidos, en lugar de los tradicionales directorios jerárquicos de carpetas y los metadatos regidos por el esquema de propiedad-valor⁴⁷⁰.

Los *topic maps* son la confluencia de tres tipos de elementos, más conocido como el TAO⁴⁷¹: los *topics* que representan los objetos descritos por el *topic map*. De forma que en un topic map acerca del CEPC⁴⁷² habrá topics tales como “constitución española”, “revistas jurídicas” o “Biglino Campo”.

⁴⁷⁰ SIGEL, A. Topic maps in knowledge organization. En PARK, J. *XML Topic Maps*. Boston : Addison-Wesley, 2003, p.445ss.

⁴⁷¹ PEPPER, S. *The TAO of Topic Maps. Finding the way in the age of infoglut*. En: <http://www.ontopia.net/topicmaps/materials/tao.html>. Consultado el 15/07/2009.

⁴⁷² Centro de Estudios Políticos y Constitucionales, sobre cuyo sitio web vamos a aplicar el modelo Topic Map para facilitar el acceso al conocimiento de sus recursos informativos a sus usuarios.

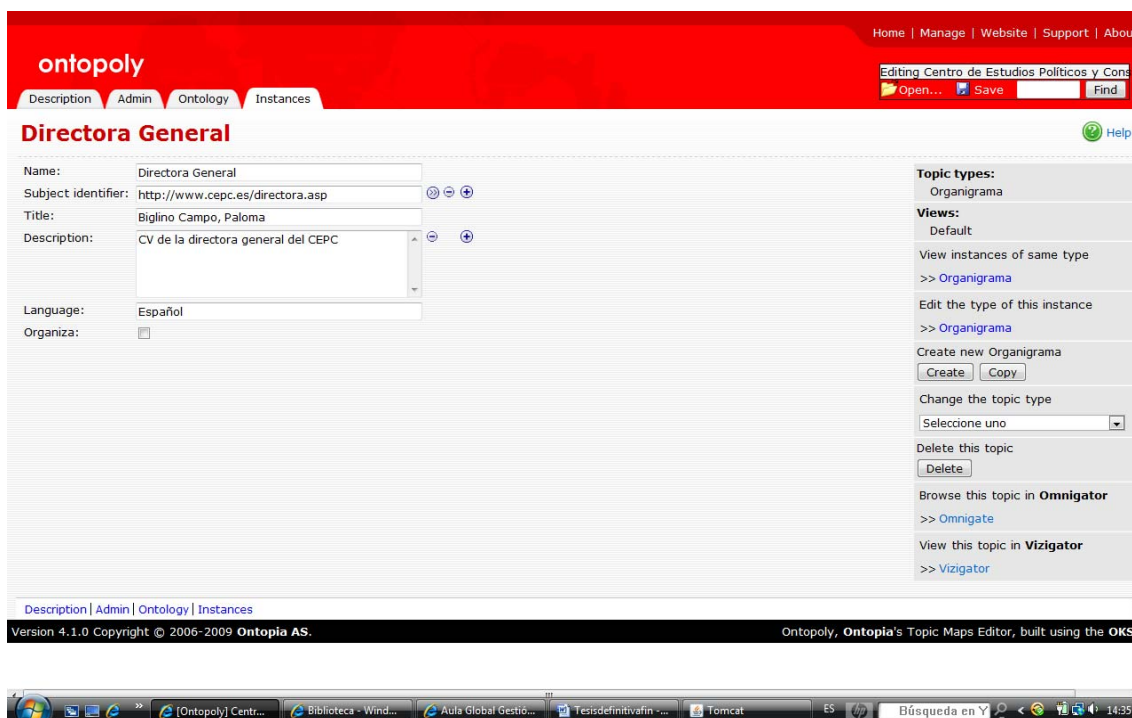


Figura 30. Muestra del *topic map* para el sitio web del CEPC mostrando el *topic* correspondiente a la directora del centro.

El segundo elemento son las *associations* que representan los distintos tipos de relación que pueden ser entabladas entre cada uno de los *topics*⁴⁷³. El tipo de *association* es expresado por una frase verbal y por el *role type* asignado a cada uno de los *topics* que participan de la *association*. Con esto se consigue que al asociar el *topic* Biglino Campo con el CEPC a través de la *association* “dirige”, los dos *topics* queden relacionados de la siguiente manera: “Biglino Campo dirige CEPC” o bien que: “el CEPC es dirigido por Biglino Campo”, ambas son la misma sentencia para el *topic map*. Es imposible declarar una sentencia sin al mismo tiempo decir la otra, de forma que las *associations* se producen en cualquiera de los sentidos. Sin embargo, las *associations* necesitan que no se les restrinja únicamente a dos *topics*. Relaciones tales como: “Biglino Campo representa al CEPC ante organismos internacionales” puede ser expresado mediante el empleo de una *association* con tres roles y no por ello es más difícil de expresar que si lo hiciéramos con *associations* más sencillas.

La última de las principales características de los *topic maps* son las *occurrences*; se trata de aquellos recursos que son pertinentes a un *topic*⁴⁷⁴. Así por ejemplo para Biglino Campo, las *occurrences* serían su página web, su CV en formato pdf, una fotografía suya en jpg, etc.

⁴⁷³ PEPPER, S. Op. Cit. p.10.

⁴⁷⁴ Ibid., p.8.

Puesto que las *occurrences* pueden ser agrupadas en tipos, se puede tipificar los distintos recursos mediante su asignación a un *occurrence type*. Esto significa que cuando un usuario entra en un *topic* y quiere más información acerca del mismo, el usuario no sólo consigue una colección de enlaces, sino que además sabe qué hace interesante cada uno de los enlaces.

Asimismo, los *topic* pueden ser categorizados por medio de otros *topics* que se denominan *topic types*⁴⁷⁵; ejemplos de ello serían los *topic type* “personal” o “institución”. Como vemos en la siguiente figura, el *topic type* es un *topic* más del *topic map*, lo cual significa que cualquiera que desee crear un *topic map* puede escoger los *topic types*, *association types*, *role types* y *occurrence types* más adecuados a su modelo de arquitectura de información. En consecuencia, el modelo es ilimitadamente extensible y adaptable a cualquier tipo de información.

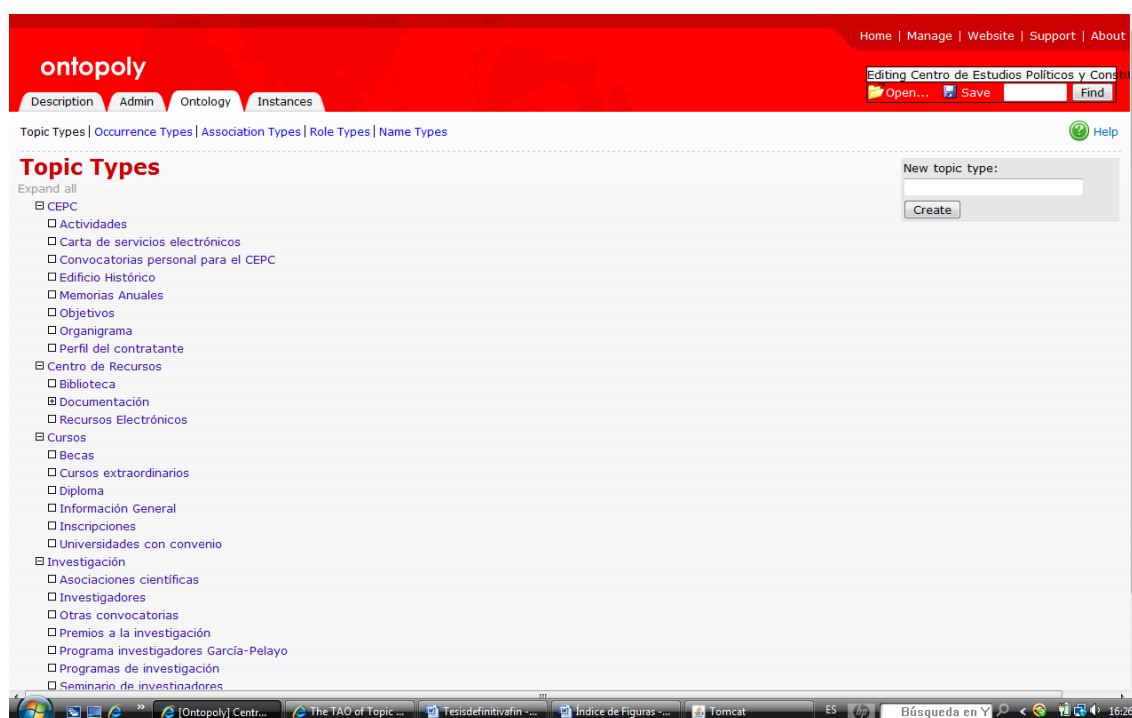


Figura 31. Relación de *topic types* para el *topic map* del CEPC.

En el diseño de un *topic map* para la gestión de una colección de recursos, se encontrará que muchos conceptos son aludidos indirectamente en la colección sin que hayan sido identificados de forma explícita. Un ejemplo, en el *topic map* para el sitio web del CEPC imaginémosnos que estamos buscando la Revista de Administración Pública, será más fácil

⁴⁷⁵ GOLDFARB, CH. ; PRESCOD, P. Op. cit. p.631.

localizarla a través del directorio de recursos del sitio que en el motor de búsqueda, puesto que nos genera más ruido en la recuperación y mucha menos precisión en los resultados pertinentes a la demanda del usuario. Con un *topic map* se dispone de un topic que representa la Revista de Administración Pública como publicación. Yendo a ese *topic* se nos presenta información, tal como, que se trata de una publicación periódica (*topic type*), o que es supervisada por un comité científico (*association*), o bien que es accesible en una URL(*occurrence*). Esto facilita notablemente las posibilidades de localización durante las búsquedas, además de mejorar la comprensión de los resultados.

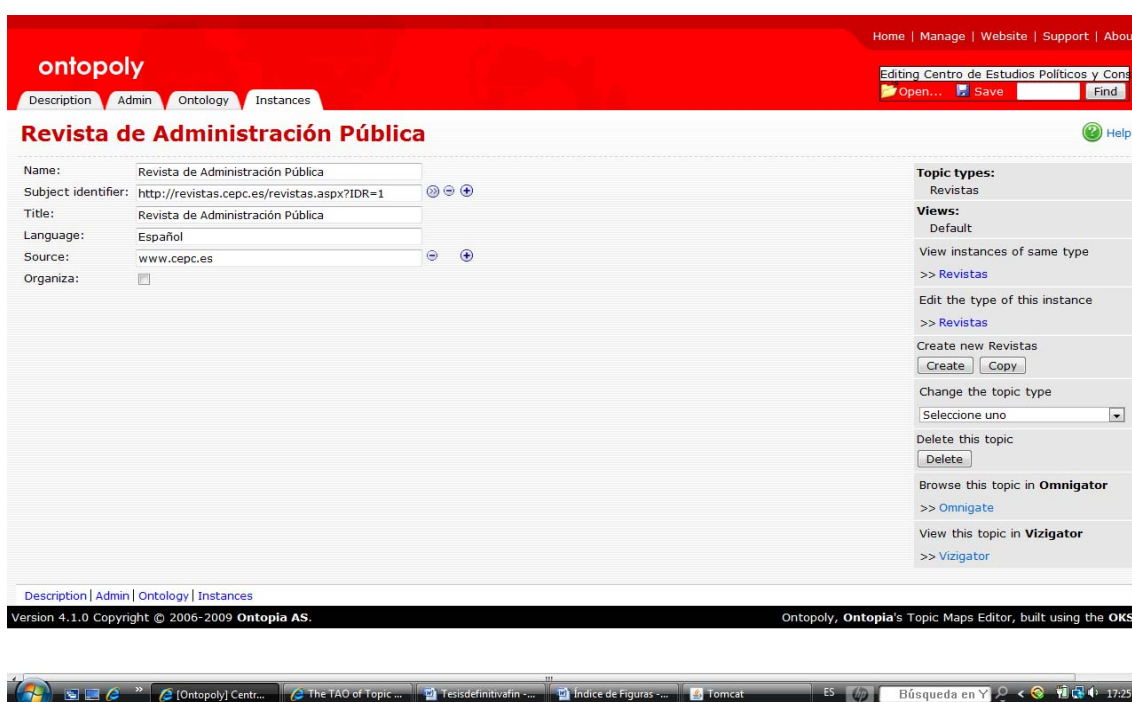


Figura 32. El *topic instance* de la Revista de Administración Pública para el *topic map* del CEPC.

La creación del topic “Revista de Administración Pública” se puede hacer bien de manera semiautomática con la introducción de los datos en la aplicación, o bien de manera manual utilizando la sintaxis XTM. En cualquier caso, las aplicaciones que construyen topic maps de forma semiautomática traducen los topics a lenguaje XTM. Así por ejemplo, la definición del topic, no sin antes asignarlo a su correspondiente topic type, que acabamos de ver se manifiesta en lenguaje XTM de la siguiente manera⁴⁷⁶:

```
<topic map xmlns="http://www.topic maps.org/xtm/1.0/"
  xmlns:xlink="http://www.w3.org/1999/xlink">
```

⁴⁷⁶ La sintaxis XTM 1.0 se puede encontrar en: <http://www.topic maps.org/xtm/>. Consultado el 15/07/2009.

```
<topic id="persona">
  <baseName>
    <baseNameString>Persona</baseNameString>
  </baseName>
</topic>

<topic id="revista">
  <baseName>
    <baseNameString>revista</baseNameString>
  </baseName>
</topic>

<topic id="Biglino Campo">
  <baseName>
    <baseNameString>Biglino Campo</baseNameString>
  </baseName>
</topic>

<topic id="Revista de Administración Pública">
  <baseName>
    <baseNameString> Revista de Administración Pública
  </baseNameString>
  </baseName>
</topic>

</topic map>
```

En la sintaxis contamos con dos topics, revista y persona, que serán topic types de nuestro topic map por corresponderse con características de los topics. Los elementos “basename” nos proporcionan la identificación asignada a nuestros dos topics, “Biglino Campo” y “Revista de Administración Pública”, que son instancias de los dos topic types precedentes, teniendo en cuenta que si bien el orden de los topic es irrelevante resulta conveniente llevar un orden jerárquico a la hora de introducirlos.

El paso siguiente es añadir un topic a ser usado como un *occurrence type*.

```
<topic id="Paloma Biglino">
  <instanceOf>
    <topicRef xlink:href="#Organigrama"/>
  </instanceOf>
  <baseName>
    <baseNameString>Biglino Campos, Paloma</baseNameString>
  </baseName>
</topic>

<topic id="REDC">
  <instanceOf>
    <topicRef xlink:href="#revista"/>
  </instanceOf>
  <baseName>
    <baseNameString>Revista Española de Derecho
    Constitucional</baseNameString>
  </baseName>
</topic>
```

```
</baseName>
</topic>

<topic id="Organigrama">
  <baseName>
    <baseNameString>Organigrama</baseNameString>
  </baseName>
</topic>

<topic id="directora general">
  <instanceOf>
    <topicRef xlink:href="#Organigrama"/>
  </instanceOf>
  <baseName>
    <baseNameString>Directora General</baseNameString>
  </baseName>
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#Directora General"/>
    </instanceOf>
    <resourceRef xlink:href="http://www.cepc.es/directora.asp"/>
  </occurrence>
</topic>
```

Los dos primeros topics correspondientes a la Revista Española de Derecho Constitucional y a Paloma Biglino Campo son instancias de los *topic types*, revista y organigrama, previamente definidos. La expresión “instanceOf” es empleada para designar el *topic* que define la clase. A continuación se define la *occurrence type* “Directora General”. El elemento *resourceRef* que se encuentra dentro de la *occurrence* proporciona la URI del recurso que constituye la *occurrence*.

Finalmente se crean *topics* para la *association* y los *role types* y se crean las correspondientes *associations* a fin de completar el *topic map*, tal como vemos a continuación:

```
<topic id="autoría">
  <baseName>
    <baseNameString>Autoría</baseNameString>
  </baseName>
</topic>

<topic id="autor">
  <baseName>
    <baseNameString>Autor</baseNameString>
  </baseName>
</topic>

<topic id="artículo">
  <baseName>
    <baseNameString>Artículo</baseNameString>
  </baseName>
</topic>
```

```
<association>
  <instanceOf>
    <topicRef xlink:href="#autoría"/>
  </instanceOf>

  <member>
    <roleSpec>
      <topicRef xlink:href="#autor"/>
    </roleSpec>
    <topicRef xlink:href="#Rodríguez Bereijo"/>
  </member>

  <member>
    <roleSpec>
      <topicRef xlink:href="#artículo"/>
    </roleSpec>
    <topicRef xlink:href="#La constitución española en su XXX
aniversario"/>
  </member>
</association>
```

Esta primera *association* es del tipo *autoría* y tiene un *topic* que realiza el papel de “autor” y otro que efectúa el papel “artículo”. La implicación de los *topics* en la *association* es descrita empleando el tipo de elemento “miembro”, y los *role type* son definidos con la expresión *roleSpec*. El *topic* participante es señalado por el “topicRef” que se encuentra dentro del “member”.

Todo esto constituye un *topic map*. Con un navegador se puede navegar por las distintas partes del mismo, hacer preguntas sobre él o bien definir una plantilla para la realización de otros *topic maps*.

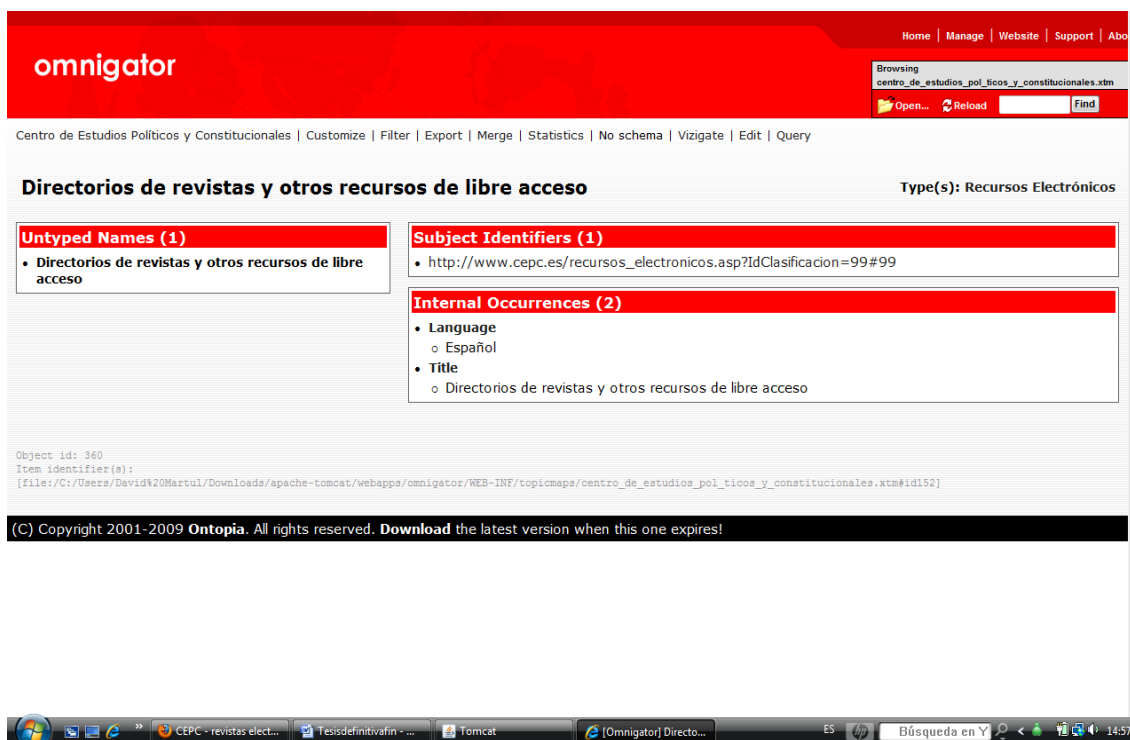


Figura 33. *Topic map* del CEPC donde se recoge el topic type del catálogo de revistas.

La cuestión siguiente sería cómo afrontar la construcción de un *topic map*. Para ello existen dos aproximaciones básicas⁴⁷⁷:

1. Que un usuario escriba manualmente el topic map, con la sintaxis XTM que acabamos de mostrar. Esto permite la obtención de topic maps de alta calidad pero a costa de un gran esfuerzo. Esto es factible para pequeños proyectos, pero inasumible para un gran proyecto.
2. La generación automática de los topic maps a partir de fuentes estructuradas de datos como XML, y aplicaciones más especializadas. Esto puede proporcionar buenos resultados si los datos existentes están bien estructurados; si no, existen distintas herramientas para el procesamiento del lenguaje natural que pueden servir de ayuda.

Para generar un topic map tan sólo se necesita un editor de texto, y para la generación automática se pueden emplear perfectamente XSLT. No obstante esta manera de generar

⁴⁷⁷ GARSHOL, L. M. *What are Topic maps*. En: http://www.xml.com/pub/a/2002/09/11/topic_maps.html Consultado el 15/07/2009.

topic maps no es suficiente cuando se trata de grandes repositorios documentales. Es por esto que se ha creado software especializado para la edición de topic maps⁴⁷⁸.

El paso siguiente es saber dónde va a actuar el topic map, es decir, dónde se va a aplicar y cómo va a ser mantenido. Para las aplicaciones más sencillas basta con el almacenamiento de documentos XTM en archivos, pero para la mayoría de las aplicaciones será necesario el almacenamiento en bases de datos.

El núcleo de una aplicación para la edición de topic maps reside en el topic map engine, equivalente al sistema gestor de los editores de bases de datos relacionales, sólo que diseñado para topic maps. Este componente sabe cómo importar y exportar documentos en sintaxis XTM y otras sintaxis específicas de topic maps como HyTime o LTM, así como almacenar, actualizar y realizar consultas en los topic maps. El motor tratará con el almacenamiento y cualquier actualización que ocurra en él. Así por ejemplo para aplicaciones que ejecuten un portal organizado con topic maps se situarán por encima del motor y lo emplearán para acceder al mismo.

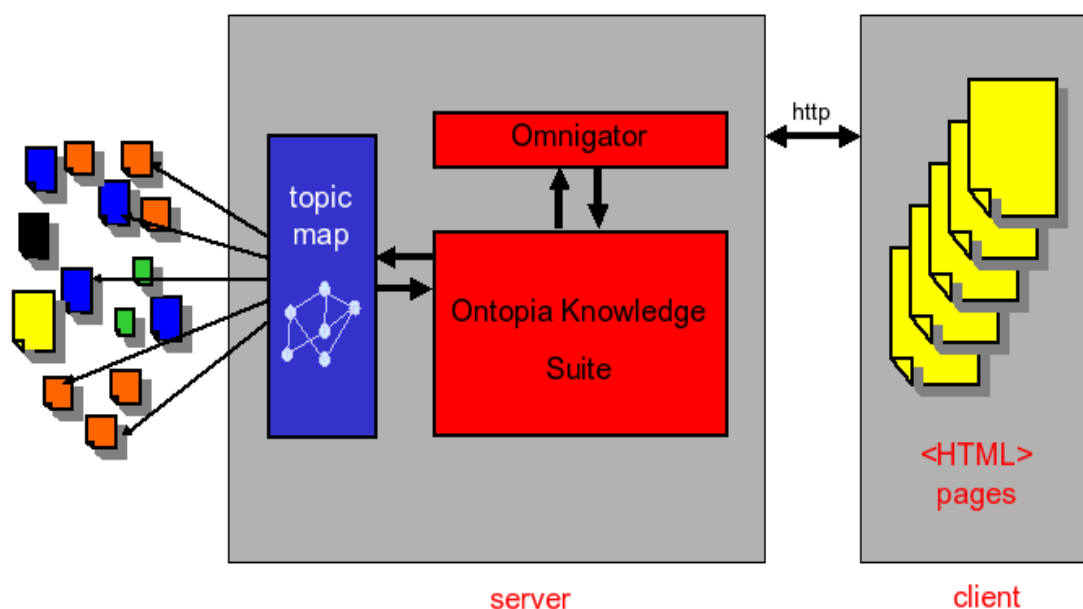


Figura 34. Arquitectura de un topic map engine: Omnigator.

⁴⁷⁸ Omnigator: The Topic Map browser. User's guide. <http://www.ontopedia.net/omnigator/docs/navigator/userguide.html>
Consultado el 21/07/2009.

Existen dos características en el modelo topic map que proporcionan un gran potencial para la recuperación al servir como filtros para precisar la localización de la información demandada por el usuario. Estos son:

I) Scope⁴⁷⁹, puede ser anexo a cualquier nombre, occurrence o association en el topic map. El scope permite no sólo la matización de una afirmación sino incluso su correcta expresión semántica. Así por ejemplo, si se hiciera un topic map acerca de las lenguas a partir de la norma ISO 639 y listas SIL de códigos de lenguas Ethnologue⁴⁸⁰; entonces deberemos ser capaces de registrar el hecho de que la norma ISO 639 asigna la lengua inglesa al grupo de lenguas germánicas, en tanto que Ethnologue la considera una lengua germánica occidental. Esto se puede representar en el topic map mediante la matización (scoping) de la asociación (association) entre el inglés y el germánico con un topic que represente la norma ISO 639, y la asociación entre inglés y germánico occidental con un topic representando Ethnologue. Análogamente, se puede emplear el scope para el registro del hecho de que el maldivo para Ethnologue se denomina Divehi para ISO 639.

De esta manera, los usuarios, pueden escoger bien ver todos los matices de los distintos recursos informativos que organiza el topic map, o bien sólo aquellos matices particulares que ligan su visión de las cosas con el modo con que desean verlo. Un uso común del matiz (scope) es mediante la creación de topic maps donde los topics tienen nombres en múltiples lenguas, permitiendo así la traducción multilingüe de los topic maps.

II) URI's para la identificación de materias⁴⁸¹. Un topic puede tener un número variable de identificadores de materia, URI's, que identifican la materia acerca de la que trata el topic. Estas URI's deben indicar los recursos que describen la materia para el usuario. Los recursos son conocidos como subject indicators, y permiten que las materias sean identificadas de forma unívoca en la web.

⁴⁷⁹ PEPPER, S. ; GRONMO, G. O. *Towards a General Theory of Scope*. En: <http://www.ontopia.net/topicmaps/materials/scope.htm>. Consultado el 15/07/2009.

⁴⁸⁰ The Ethnologue: Languages of the World (El etnólogo: lenguajes del mundo) es una publicación impresa y virtual de SIL International (antes conocida como el Instituto Lingüístico de Verano), una organización de servicios lingüísticos que estudia los lenguajes menos conocidos para proveer servicios a sus hablantes. Posee estadísticas de 6.912 idiomas, incluyendo el número de hablantes, su ubicación, dialectos, afiliación lingüística, disponibilidad de textos, etc. Actualmente es uno de los inventarios más comprensivos de los lenguajes existentes, junto al Linguasphere Register.

⁴⁸¹ PEPPER, S. ; SCHWAB, S. *Curing the Web's Identity Crisis*. En: <http://www.ontopia.net/topicmaps/materials/identitycrisis.html>. Consultado el 15/07/2009

Esta identificación desambigua de las materias se emplea en los topic maps para intercalar topics de la misma materia. Básicamente, lo que ocurre, es que dos topics que designan la misma materia son reemplazados por un topic que presenta la unión de las características de ambos topics (nombres, occurrences y associations). Sin embargo, ¿cómo podemos conocer la identidad de una materia? Como veremos en el siguiente capítulo se hará a través de los PSI's publicados por parte del consorcio OASIS, dada su dedicación a la creación de normas procedimentales sobre la creación, publicación y mantenimiento de indicadores de materia para su empleo generalizado en la web. Un ejemplo de ello es el empleo de URI's para todos los países del mundo, basado en la norma ISO 3166 para códigos de países, que permite la inferir el mismo significado cuando en un topic map se diga España y en otro español⁴⁸².

Los topic maps hacen que la información sea localizable mediante la asignación de una identidad a cada concepto que aparezca en un recurso de información, proporcionando así múltiples estrategias de navegación. Estas estrategias son semánticas, es decir, todas las identidades son identificadas con nombres y tipos que le indican al usuario desde el propio corpus donde se encuentra, lo que vendría a constituirse en lo que Goldfarb señalaba acerca de los topic maps⁴⁸³ como “el GPS del universo de la información”.

Estas dos propiedades de los topic maps no sirven de mucho si no somos capaces de insertarlas adecuadamente en una arquitectura de información capaz de relacionar la organización del conocimiento con los recursos de información de un dominio particular. Cualquier trabajo de arquitectura necesita de algún tipo de pasarela si deseamos, como documentalistas, proporcionar al usuario un acceso eficaz y eficiente a los recursos que deseamos poner a su alcance. Para ello, la arquitectura debe ser capaz de hacer interoperables los distintos sistemas de tratamiento documental tanto de nuestro centro como de otros centros con los que podamos actuar en red en la organización y recuperación tanto de recursos de información, como de objetos de aprendizaje o esquemas de análisis de contenido.

⁴⁸² OASIS. En: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=geolang Consultado el 15/07/2009

⁴⁸³ GOLDFARB, Ch. Topic maps: Knowledge navigation aids. En GOLDFARB, Ch. *Charles Goldfarb's XML handbook*. New Jersey : Prentice Hall, 2002. p.645.

En la mayoría de las organizaciones, la arquitectura de información trata con un rango de recursos heterogéneos: documentos digitales en distintos formatos, bases de datos, canales de noticias, así como documentos en papel. Para poder proporcionar acceso a cada uno de estos recursos la solución que se ha venido proporcionando ha sido proporcionar un interfaz de búsqueda o bien de navegación para cada uno de estos formatos. Esto ha supuesto una multiplicidad de portales y pasarelas que dificultan un acceso unívoco a un cluster de recursos correspondientes a un dominio de interés por parte del usuario.

Hasta el momento, la respuesta a este problema se ha dado en el campo de los motores de búsqueda web, al contar con la opción de búsqueda por materias, que junto al empleo de filtros, facilitan búsquedas en ámbitos más acotados. Sin embargo, los resultados de las búsquedas no producen resultados semánticamente pertinentes, dado que para un motor una estrategia de búsqueda no es más que una sucesión de signos sin ningún tipo de significación. Así, si deseamos buscar en Google el concepto de “Derecho”, nos ofrece más de 92.000.000 de enlaces⁴⁸⁴ a páginas con la palabra “Derecho”, pero muchos de estos documentos no tienen nada que ver con la materia. En el catálogo⁴⁸⁵ del CEPC, realizando una búsqueda por “Derecho” en el campo de materias, nos ofrece 27287 registros que por supuesto tratan de algún aspecto del Derecho, pero ninguno de los 10 primeros resultados ofrece alguna obra de esta materia como concepto o disciplina.

Como vemos los motores y los catálogos⁴⁸⁶ están más diseñados para una recuperación exhaustiva que para obtener resultados precisos, menos por contexto y mucho menos por la semántica de los términos. Son por tanto herramientas de recuperación muy dependientes de los recursos de la colección de la que dependen.

Otro de sus defectos es que la recuperación la realizan siempre sobre recursos textuales por los que se pueda realizar la indización. Esto es incoherente con la progresiva inserción en la red de documentos multimedia, como por ejemplo en revistas digitales, o el sitio web del CEPC, incapaz de recuperar cualquier tipo de imagen y mucho menos documentos multimedia u objetos de aprendizaje propios para los cursos que allí se imparten.

⁴⁸⁴ Catálogo del CEPC: Derecho. Consultado el 10/07/2009.

⁴⁸⁵ Catálogo Absys del CEPC. Consultado el 10/07/2009.

⁴⁸⁶ Si bien es cierto que hoy en día los catálogos incorporan herramientas de asistencia al usuario como las listas de autoridad del catálogo del CEPC o los lenguajes documentales que sugieren al usuario los términos por los que realizar las búsquedas.

En entornos institucionales, como el CEPC, la combinación de los motores de búsqueda con las taxonomías⁴⁸⁷ y directorios ha venido significando una apuesta por el empleo combinado de las estrategias de búsqueda analítica y las estrategias de búsqueda por navegación⁴⁸⁸. Resultado de lo cual ha sido la aparición de los portales institucionales que ofrecen mapas de sitios y motores de búsqueda propios para la recuperación en el propio sitio web. Sin embargo, estas herramientas de asistencia al usuario no son suficientes si no se integran en un marco de gestión del conocimiento en el entorno corporativo. Es decir, las instituciones demandan sistemas capaces de gestionar todas las etapas del ciclo documental bien en los portales propios o bien en los vortales⁴⁸⁹ de su ámbito de especialidad.

Fruto de esta demanda ha sido la aparición de los topic maps, tecnología web construida en lenguaje XML y cuya especificación, de acceso libre, ofrece un medio menos complejo, más barato y más interoperable para proporcionar acceso a una colección de recursos de información heterogéneos, con la ventaja adicional de fundamentarse simultáneamente en los principios de los lenguajes precoordinados como las clasificaciones y en los lenguajes postcoordinados como los tesauros⁴⁹⁰.

Aunque el término topic map es poco conocido en el ámbito académico, la funcionalidad a la que se refiere no lo es tanto. Un topic map es esencialmente un índice al contenido de una colección de recursos informativos particulares, tanto de nuestro portal corporativo como de un ámbito de conocimiento especializado o vortal. Debido a que son creados en soporte electrónico, presentan una serie de propiedades que los diferencian de los índices alfabéticos en soporte papel, sobre todo en cuanto a la dimensión de los espacios de información. Nos referimos con esto al hecho de que el documento en papel nos presenta la realidad en dos dimensiones, obligándonos a una lectura lineal y jerárquica, con pocas posibilidades para la interrelación por parte del lector con otros apartados del texto que no

⁴⁸⁷ Es significativo de esto el trabajo que el profesor Alan Gilchrist realiza en la empresa Cura Consortium, dedicada a la resolución de problemas de gestión de la información en las empresas.

⁴⁸⁸ MARCHIONINI, G. *Information seeking in electronic environments*. Cambridge: Cambridge University Press, 1997.

⁴⁸⁹ Tal y como recoge García Gómez, un portal corporativo abarca en sentido vertical y jerárquico la totalidad de recursos de la colección. En tanto que el vortal tiene un sentido más transversal al recoger los recursos informativos relacionados con un dominio concreto de conocimiento. En GARCÍA GÓMEZ, J.C. Portales de Internet: concepto, tipología básica y desarrollo. *El Profesional de la Información*, julio-agosto, vol.10, n° 7-8. 2001. pp.4-13.

⁴⁹⁰ SLYPE, Georges van. *Lenguajes de indización : concepción, construcción y utilización en los sistemas documentales*. Madrid : Fundación Germán Sánchez Ruipérez , 1991.

sean adyacentes a lo que está siendo leído. Los textos digitales nos permiten en cambio la navegación, la intertextualidad, saltar e interrelacionar libremente los distintos apartados del texto sin necesidad de seguir una jerarquía de epígrafes para comprender la globalidad del texto. Son documentos en tres dimensiones, y la tercera dimensión la aportan los hiperenlaces bien a otras secciones de texto o bien a otros textos que no necesariamente tienen que estar físicamente con el lector⁴⁹¹.

En Internet, se hace evidente que el modelo de lectura lineal y jerárquica ha sido abandonado en favor de una red definida por una alta densidad de relaciones. Incluso antes de la llegada de la web ya se habían estado realizando iniciativas para escapar de la restricción bidimensional del formato papel. Entre los primeros en abordar este problema se encontraban los pioneros de la inteligencia artificial y los teóricos de la psicología cognitiva cuyos modelos multidimensionales de adquisición del conocimiento pretendían emular el pensamiento asociativo del hombre en los ordenadores⁴⁹².

Una de estas iniciativas fue PRECIS⁴⁹³ (Sistema de indización con contexto conservado), desarrollado por Derek Austin a comienzos de los años 70. Este sistema aplicaba el poder de procesamiento de los ordenadores para la producción de un sofisticado formato de índice permutado que preservaba el contexto sintáctico y semántico de la palabra clave. Fue utilizado por primera vez por la Bibliografía Nacional Británica en 1971, pero a pesar de ser suplantado por el simplificado COMPASS, fue finalmente abandonado en 1996 debido a su extrema complejidad para ser manejado manualmente. Sin embargo, el problema no residía en sí mismo en PRECIS sino en el hecho de aplicar la capacidad de procesamiento computacional al espacio de información inherentemente limitado de las páginas impresas⁴⁹⁴.

El problema con la información en formato papel es que no permite la conservación efectiva del contexto. Ellis a finales de los 80 afirmaba que los sistemas de recuperación de

⁴⁹¹ BORGMAN, Ch. *From Gutenberg to the Global Information Infrastructure : access to information in the networked World*. Cambridge: MIT Press, 2000.

⁴⁹² BATER, B. Topic maps. Indexing in 3-D. En: Gilchrist, A. ; Mahon, B. *Information Architecture*. London : Facet Publishing, 2004. p.133.

⁴⁹³ AUSTIN, D. *PRECIS: a manual of concept analysis and subject indexing*. London : Council of the British National Bibliography, 1974.

⁴⁹⁴ BATER, B. Op. Cit., p.134.

información deben proporcionar más canales para la navegación⁴⁹⁵. Belkin y Marchetti fueron incluso más lejos al defender que estos sistemas debían ser diseñados empleando una red de asociaciones entre palabras clave como medio para cubrir los vacíos semánticos. Defendían que cambiando la pregunta del “¿Qué es una demanda informativa?” por “¿Por qué la información necesita existir?”, la noción de contexto podría ser abordada como característica en los sistemas de recuperación⁴⁹⁶.

4.4.3 Concepción holística del modelo *Topic map*.

El potencial holístico⁴⁹⁷ de los topic maps está en que es una herramienta para la navegación que combina una estructura articulada de topics interconectados y navegables con un interfaz flexible para proporcionar un entorno contextual más rico al usuario⁴⁹⁸. La navegación es un concepto inherentemente dinámico, lo que hace difícil su aplicación a documentos en un formato físico como el papel, tanto con el empleo de índices como de cualquier otro lenguaje documental, pero en el ámbito de los hipermedios es una estrategia de búsqueda alternativa a la proporcionada por los motores de búsqueda. Sin embargo, para que la búsqueda por navegación sea efectiva, y el usuario sea capaz de resolver sus demandas informativas de forma satisfactoria y en un breve periodo de tiempo necesita contar con referentes que le ayuden a orientar la navegación. Estos referentes los proporciona un topic map, un mapa de materias con el que poder representar virtualmente cualquier tipo de estructura de conocimiento o modelo de datos y guiar las búsquedas del usuario⁴⁹⁹. En este sentido, con los topic maps, la búsqueda de documentos y de conocimiento por navegación se convierte en una opción alternativa seria a los sistemas de recuperación de información tradicionales, consolidando la navegación como herramienta para la gestión del conocimiento en la arquitectura de información de cualquier entorno digital.

⁴⁹⁵ ELLIS, D. A behavioural approach to information retrieval design. *Journal of Documentation*, vol. 45, n° 3, 1989, p.173.

⁴⁹⁶ BELKIN, N. J. ; MARCHETTI, P. BRAQUE: design of an interface to support user interaction in information retrieval. *Information Processing and Management*, vol. 29, n°3, 1993, p. 331ss.

⁴⁹⁷ Entendemos por holístico tal y como lo expresa el diccionario de la RAE. Holismo: Doctrina que propugna la concepción de cada realidad como un todo distinto de la suma de las partes que lo componen. Es decir, deseamos proporcionar una visión global y práctica de lo que son los topic maps con independencia de la descripción analítica que proporcionamos en el capítulo siguiente.

⁴⁹⁸ BATER, B. Op. cit., p.135.

⁴⁹⁹ PEPPER, S. Topic Maps. En BATES, M. ; MAACK, M. *Encyclopedia of Library and Information Sciences*. Boca Raton, FL : CRC Press, 2009.

Entre las razones por las cuales los topic maps no se han implantado en la gestión del conocimiento en entornos digitales reside en el hecho de que la mayoría de ellos se aplican a un ámbito relativamente local como si se trataran de bases de datos locales, cuando sus inicios están en entornos de información masivos y globales, para sistemas distribuidos y semánticamente integrados⁵⁰⁰. No obstante, se está produciendo un incremento progresivo de sistemas topic map derivados de su uso en los sistemas de información, que con el uso incipiente de la web social, se vuelven más necesarios. Así, han pasado en un breve periodo de tiempo, de organizar la información personal a la creación y gestión de sitios web. Propósito que planteamos en esta tesis para la gestión del sitio web del CEPC como una biblioteca digital semántica.

Algunos casos prácticos de aplicación de topic maps para la gestión de entornos digitales son:

i) Mondeca⁵⁰¹ empresa para la que trabaja uno de los primeros teóricos de los topic maps Bernard Vatant, ha desarrollado sistemas gestores como ITM T3, un servidor para el trabajo colaborativo, basado en el contexto para la creación, mantenimiento y distribución de taxonomías, terminologías y tesauros que proporcionen a las empresas soluciones de gestión de información sobre la base de la aplicación de tecnologías semánticas que ayuden a las organizaciones a obtener el máximo rendimiento de su conocimiento.

ITM T3 se basa en una plataforma semántica topic map para la gestión de topics, Mondeca Topic Navigator ya en uso por parte de organizaciones como Thomson, LexisNexis, Lafarge y Wolters Kluwer⁵⁰². ITM T3 permite a las empresas la gestión de vocabularios de referencia y que éstos sean interoperables con otras aplicaciones del sistema de información como motores de búsqueda, herramientas de minería de texto, traducción y clasificación.

ii) Se emplea un sistema topic map en el laboratorio nacional de Oak Ridge⁵⁰³ como parte de un sistema de asistencia a los oficiales en la aplicación de reglas de clasificación de

⁵⁰⁰ PASSIN, T. *Explorer's Guide to the Semantic Web*. Greenwich : Manning, 2004, p. 87.

⁵⁰¹ Mondeca. En: www.mondeca.com. Consultado el 15/07/2009.

⁵⁰² ITM T3. En: http://mondeca.com/index.php/fr/news/archives/mondeca_announce_itm_t3 Consultado el 15/07/2009.

⁵⁰³ OakRidge National Laboratory. En: <http://www.ornl.gov/> Consultado el 15/07/2009.

seguridad de los documentos allí depositados. El topic map gestiona tanto las reglas de seguridad como los documentos, y un motor de inferencia aplica las reglas a los datos en el topic map.

iii) El servicio estadounidense para la tributación de rentas internas lanzó un CD que contiene todas las regulaciones y documentos actualizados para su pago. Todo ello indizado a partir de un topic map⁵⁰⁴ diseñado por uno de los teóricos de los topic maps, Biezunski. Actualmente la aplicación es accesible por internet.

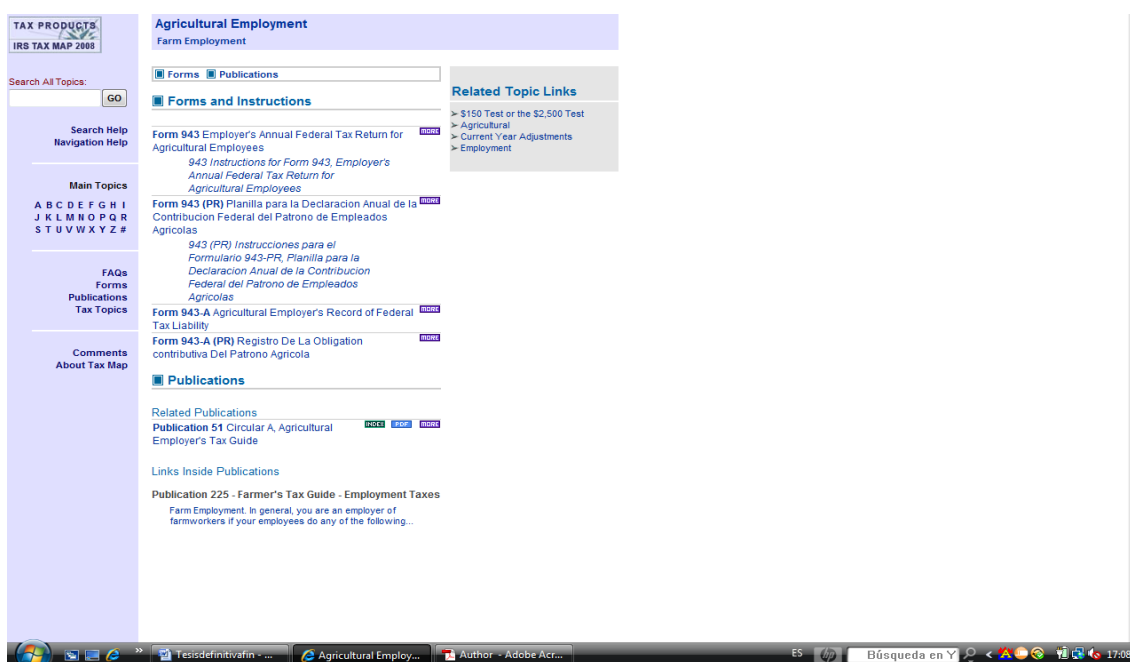


Figura 35. Ejemplo de búsqueda de información tributaria con el Tax Map.

iv) La empresa Borland International creó un sistema que integra distintas bases de datos añadiéndoles un software que las hace actuar como un único topic map⁵⁰⁵. El topic map virtual se encuentra en la red de la compañía. Esta aplicación topic map comienza a aproximar a la visión de hiperenlaces y navegación por un abanico de diferentes documentos distribuidos por la red.

v) LexisNexis, que mantiene para sus suscriptores una gran colección de referencias en línea, ha desarrollado un sistema topic map que facilita a sus usuarios la localización de la

⁵⁰⁴ IRS Tax Map. En: <http://taxmap.ntis.gov/taxmap/>. Consultado el 15/07/2009.

⁵⁰⁵ SEIDL, F. Elmer Preview: A standards-based architecture for creating a unified information space. En *XML 2001*. En: www.idealliance.org/papers. Consultado el 15/07/2009.

información de su interés⁵⁰⁶. Este sistema destaca por no almacenar las referencias en el topic map, sino que el topic map contiene metadatos que son empleados para la identificación y localización de información en los principales repositorios de la empresa.

vi) La administración estadounidense de la seguridad social ha desarrollado un sistema gestor de contenidos para documentos relacionados con las pólizas, a los que se accede por medio de un topic map de todas las empresas de Estados Unidos que cotizan a la seguridad social⁵⁰⁷. Para ello, se empleó un primer trabajo que recogía las referencias de la base de datos de MEDICARE para usuarios corporativos. Esta iniciativa surgió por la necesidad de tratar con mayor rapidez los casos de apelaciones médicas complejas. El sistema permitía a las empresas indizar sus propios documentos empleando topics comunes relacionados con los procedimientos y terminología de la aseguradora MEDICARE. Los expertos en materia de pólizas añaden y mantienen las palabras clave en el topic map sin tener que depender del personal técnico. Esto ayuda a conservar los topics con los que están familiarizados las personas que solicitan información, y mejora la capacidad del usuario final para la localización de información compleja e interrelacionada.

El software empleado para el diseño de todos estos topic maps es muy variado, pero fundamentalmente podemos distinguir tres tipos de software⁵⁰⁸:

- a) Software autónomo, disponible comercialmente,
- b) otros son contruidos a partir del empleo de bases de datos ya existentes, y
- c) algunos son proyectos con código libre para el diseño de topic maps disponibles para proyectos de pequeña escala.

En todos estos casos los topic maps ofrecen una capacidad sin precedentes para la realización de búsquedas en general, y especialmente de búsquedas a texto completo, dado su potencial para la integración de consultas en lenguaje natural. Las búsquedas a texto

⁵⁰⁶ FREESE, E. Taking Topic Maps to the Nth dimensión. En *Extreme Markup Languages* 2003. En: www.mulberrytech.com. Consultado el 15/07/2009.

⁵⁰⁷ DEGLER, D. ; BATTLE, L. Can Topic Maps describe context for Enterprise-wide applications? En *Extreme Markup Languages* 2003. En: www.mulberrytech.com/ Consultado el 15/07/2009.

⁵⁰⁸ PASSIN, T., Op. cit., p.88.

completo en un corpus documental consisten en la recuperación de documentos a partir de un término que aparece en el texto de uno de los documentos. Sin embargo, en un sistema basado en topic maps, en vez de recuperar los documentos se recuperan los topics⁵⁰⁹ que más se adecúan a la temática del término buscado. Así, es posible proporcionar un punto de partida para orientar la navegación en la consecución de una respuesta a la consulta efectuada.

En nuestra propuesta de biblioteca digital semántica para el CEPC, una vez aplicada la arquitectura de sitio mediante un topic map, podríamos utilizar un lenguaje de consulta para el topic map⁵¹⁰ que trabaje de la siguiente manera. Si un usuario desea recuperar el capítulo de la obra de Solé Tura⁵¹¹ sobre la constitución de 1812 a partir del topic map lo puede hacer de forma poliédrica. En esta obra, un referente en la historia contemporánea y en ciencias políticas, se comentan los distintos factores sociopolíticos que intervinieron en la génesis de las constituciones españolas desde 1808 hasta 1936. Existe un capítulo destinado a la constitución de 1812, de donde es posible recuperar las circunstancias políticas, sociales, así como una bibliografía muy completa acerca de la temática. A través del topic map es posible recuperar todos aquellos documentos que tratan sobre las distintas materias o topics relacionados con la constitución de 1812, de forma que es posible un modo de recuperación poliédrico de los distintos epígrafes que constituyen este capítulo. Por medio de un lenguaje de consulta, específico de los topic maps⁵¹², es posible la formulación de consultas tan complejas como el usuario desee.

El principal problema en la recuperación, si aplicamos topic maps, es que el usuario necesita conocer tanto su lenguaje de consulta TMQL como el propio topic map que rige la arquitectura de información de la biblioteca digital del CEPC. Normalmente los usuarios tienen grandes carencias en habilidades para las tecnologías de la información y la comunicación y muchas más en el manejo de los lenguajes documentales y los topic maps. Como veremos este problema puede ser suplido con la aplicación a nuestro topic map de un interfaz de usuario usable que permita la elaboración de sentencias de búsqueda por

⁵⁰⁹ E información adicional como las occurrences o las associations que ayudan a fijar el contexto semántico del término.

⁵¹⁰ TMQL es un standard del que hablaremos más adelante.

⁵¹¹ SOLÉ TURA, J. *Constituciones y periodos constituyentes en España (1808-1936)*. Madrid : Siglo XXI, 1990.

⁵¹² La norma ISO 18048 es la que se refiere a TMQL. GARSHOL, L ; BARTA, R. Topic Maps Query Language. En: <http://kill.dev.at/system/files/tmq.pdf> Consultado el 15/07/2009.

medio de la selección de términos a partir de una lista integrada en el interfaz gráfico de usuarios denominada drop-down list⁵¹³, tal y como vemos en la siguiente figura.

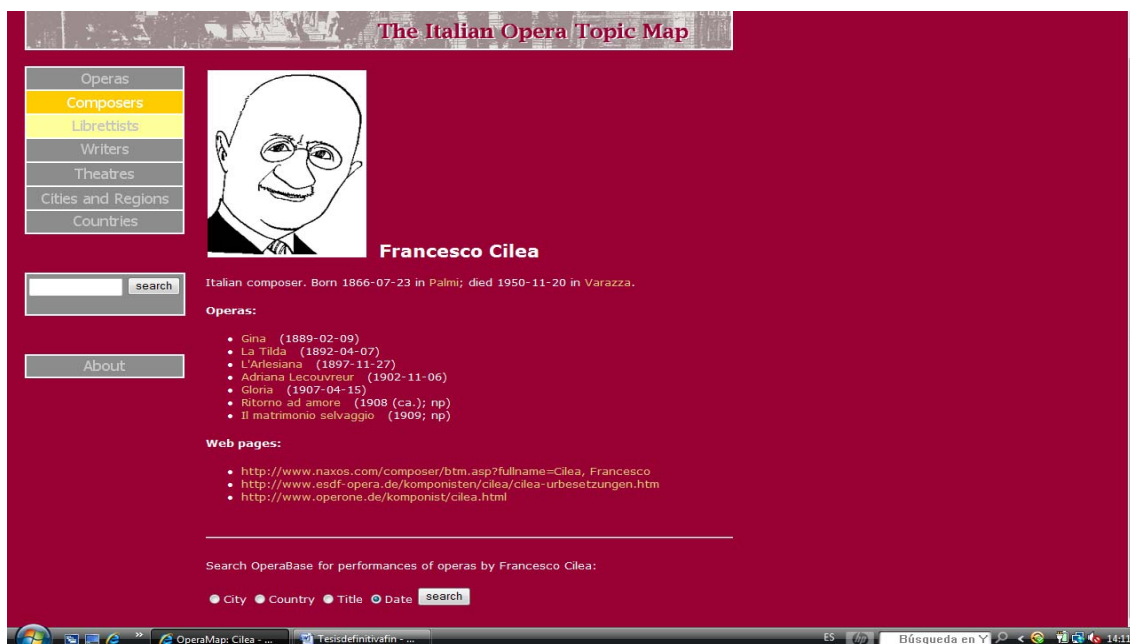


Figura 36. Ejemplo de interfaz gráfico drop-down list para un topic map.

En los últimos años se han puesto a disposición de los usuarios una serie de herramientas para que puedan construir sus propios topic maps⁵¹⁴. Esto permite que los usuarios puedan definir un vocabulario, bien en lenguaje libre como especializado, para la representación personalizada de un dominio de conocimiento. Sin embargo, pueden incurrir en un incorrecto diseño del topic map si no definen previamente un conjunto de reglas que doten al topic map de coherencia. Así por ejemplo, puede que empleen las associations de forma que carezca de sentido adjuntar occurrences a los topics integrados en un topic type inadecuado.

Para las bases de datos relacionales, modelo de Chen, y XML la solución tradicional a este problema es la definición de un schema que recoja tanto la estructura formal permitida para una estructura de datos particular como los elementos básicos a tener en cuenta. Esta misma solución es la que se proporciona en los topic maps con el empleo del lenguaje

⁵¹³ La drop-down list es un dispositivo de control de la interfaz gráfica de usuario que permite la selección múltiple de términos de entre una lista de opciones. PARK, J. *XML Topic Maps. Creating and using Topic Maps for the Web*. Boston : Addison-Wesley, 2003. pp.271-273.

⁵¹⁴ Es el caso de Omnigator, que acaba de pasar a ser software libre. En: <http://code.google.com/p/ontopia/downloads/list> Consultado el 15/07/2009.

TMCL⁵¹⁵ que proporciona un marco estable para su diseño. Por medio de este lenguaje para schemas es posible efectuar declaraciones del tipo: “sólo las obras pueden ser creadas por uno o más autores”. Esto, además, nos permite identificar las potencialidades de los topic maps con una sintaxis propia para la definición de *schemas* con el cual describir reglas a seguir por el software de edición de topic maps.

Los topic maps cuentan con un mecanismo para la declaración formal de identidad de la materia de un topic⁵¹⁶. Esto significa que en los topic maps existe un modo de declaración de la materia de un topic, de forma que si se localiza otro topic sobre la misma materia, se pueda asegurar con certeza que ambos topics son representación de la misma materia.

El mecanismo de identificación en topic maps se basa en las URIs, las cuales pueden ser empleadas de dos maneras diferentes. Si la materia del topic es un recurso de información, entonces la materia puede ser identificada muy fácilmente dado que tan sólo es necesario referirse al recurso por el empleo de una URI. De este modo dos topics que se refieran a un mismo objeto o recurso de información de esta manera, deben ineludiblemente representar de forma unitaria el mismo recurso con un único topic.

El problema se complica cuando la materia no puede ser representada por medio de un recurso de información dado que se trata de una materia abstracta. Por ejemplo, la constitución de 1812 es una materia concreta fácilmente representable por medio de un hiperenlace al texto de esta constitución, pero ¿cómo podemos representar el concepto o materia de la soberanía nacional? Esta última materia carece de una fácil definición por medio de una URI. Lo que se ha hecho es permitir la creación de un documento en el que se defina la materia para que a través de una URI se pueda apuntar a este recurso definitorio.

No obstante esto puede llevar a confusión dado que no está claro si la materia es el recurso al que está referido o bien el concepto descrito por el recurso. Los topic maps resuelven este problema mediante la distinción entre ambos tipos de referencias para que siempre quede clara la semántica del contenido con independencia del tipo de fuente de la que se

⁵¹⁵ TMCL. *Topic Map Constraint Language (ISO19756)*. En: <http://www.isotopic-maps.org/tmcl/>. Consultado el 15/07/2009.

⁵¹⁶ PEPPER, S. ; SCHWAB, S. *Curing the Web's Identity Crisis*. En: <http://www.ontopia.net/topic-maps/materials/identitycrisis.html#N421> Consultado el 15/07/2009.

trate. La resolución del problema es por medio de la unificación (merging). La unificación consiste en que si localizamos dos topics en topic maps diferentes y de los cuales se pueda realizar una misma afirmación del estilo: “tal materia se encuentra definida por este recurso informativo”, entonces ambos topics son la representación de la misma materia⁵¹⁷.

La capacidad para conocer cuándo dos topics representan un mismo recurso u objeto informativo, proporciona a los lenguajes documentales de una gran capacidad para la ordenación de los recursos digitales con independencia de dónde éstos se encuentren, de su formato y del tipo de fuente de información de la que se trate. Este potencial puede ser explotado de las siguientes maneras⁵¹⁸:

En primer lugar permite que aquellos topic maps, contruidos en principio a partir de distintos recursos, puedan ser unificados automáticamente siempre y cuando se produzca una superposición de recursos⁵¹⁹.

En segundo lugar aquellos módulos topic map que contengan información utilizada conjuntamente pueden ser conservados separadamente y reutilizados en distintos contextos cada vez que sean necesarios.

En tercer lugar los vocabularios comunes pueden ser desarrollados como colecciones de topics definidos y reutilizados para distintos repositorios documentales, de forma que la información conservada en distintos centros de documentación pueda ser unificada o intercambiada como desee cada uno de sus usuarios.

En cuarto lugar la información obtenida en distintos recursos o documentos o repositorios puede ser presentada en un todo significativo, el topic map.

Finalmente la unificación de topic maps permite que distintos sitios web, con una arquitectura construida a partir de topic maps, puedan ser integrados en un único sitio web con una arquitectura unitaria, dado que los topics que tratan acerca de las mismas materias

⁵¹⁷ GARSHOL, L. *Metadata? Thesauri? Taxonomies Topic Maps!* En: http://www.ontopia.net/topic_maps/materials/tm-vs-thesauri.html#N1051 Consultado el 15/07/2009

⁵¹⁸ Id.

⁵¹⁹ La superposición de datos se produce cuando dos recursos pueden ser representados por una misma URI.

pueden ser fusionados en un solo topic. Esta propiedad de los topic maps es la que más utilizaremos para el diseño del topic map del CEPC.

Los topic maps proporcionan un índice orientado a los contenidos de una colección de documentos, de estructura análoga al índice alfabético de un libro pero con la diferencia de que el índice del libro no describe los contenidos del libro como una colección de materias interrelacionados, sino como una colección de referencias a materias casi siempre aisladas con referencias cruzadas a otras materias de manera ocasional.

Actúa como una colección de materias interrelacionadas con el propósito de indizar una colección documental; de forma que, con varios topic maps, es posible indizar una misma colección de documentos digitales, al igual que un libro puede tener distintos índices tales como índices de autores o de materias. Lo importante es que se pueda contar con múltiples topic maps para designar de distintas maneras las materias. De esta manera, pueden ser concebidos como una superposición de capas de información sobre documentos digitales, sobre los cuales es posible la navegación por contenidos con independencia del tipo documental del que se trate. De esta manera, actúan como taxonomías en tanto que describen, clasifican, e indizan un espacio de información integrado por documentos⁵²⁰.

Una manera de entender los topic maps en relación con los esquemas de clasificación tradicionales es como una extensión de los lenguajes documentales tradicionales pero a un nivel más elevado. Esto es, los tesauros extienden las taxonomías por medio de la adición de un sistema de propiedades y relaciones entre los nodos. Los topic maps no añaden ningún vocabulario predeterminado sino que proporcionan un modelo suficientemente flexible como para manejar un vocabulario abierto.

El primer autor que se planteó la representación de los lenguajes documentales tradicionales por medio de topic maps fue Kal Ahmed, quien definió un topic map normalizado para la representación de tesauros con topic maps. Esencialmente existen dos maneras de hacerlo: a) aquella que modela los términos como materias, un *topic* por término; b) aquella que modela los términos como nombres, un nombre por término o una materia por *topic*. En esta última aproximación se estaría creando un *topic map* donde la materia de cada *topic* vendría expresada por un nombre de término.

⁵²⁰ GARSHOL, L. Op. cit.

Vamos a emplear la primera aproximación para generar un topic map a partir de un tesauro ya definido por tratarse el dominio del derecho de un lenguaje de especialidad bien delimitado. La segunda aproximación la haríamos para aquellos campos con un lenguaje de especialidad todavía no constituido o mal definido.

Ahmed nos indica algunos principios a seguir en el caso de la construcción de un topic map a partir de un tesauro como son: 1) Para cada uno de los términos aceptados o descriptores crear un topic de tipo “término” y asignarle el nombre del término en cuestión, 2) Para cada uno de los términos no aceptados, que son parte de la relación de equivalencia “usado por (UP)” asignar a un topic con nombre de uno de los términos aceptados el nombre del término no aceptado, sólo que esta denominación del topic se encontrará en el ámbito o scope de los términos no aceptados, 3) una nota de alcance es una *occurrence* del tipo “nota de alcance”, 4) La relación RT (término relacionado) es representada como una relación asociativa de término relacionado en el topic map, 5) las relaciones jerárquicas son representadas como relaciones del tipo “genérico/específico”, donde los roles especifican qué topic es genérico y cuál es específico⁵²¹.

Para las clasificaciones facetadas la situación es muy similar. Existe una norma topic map para clasificaciones facetadas basada en XFML, así como también una norma de hoja de estilo XSLT para la conversión de documentos XFML a topic maps⁵²². Las clasificaciones facetadas en XFML extienden las capacidades de los tesauros mediante la adición del tipo de topic “faceta” y un nuevo tipo de association “pertenece a la faceta”.

La creación de un *topic map* a partir de un sistema de clasificación facetada puede realizarse siguiendo las siguientes etapas⁵²³: 1) para cada faceta se crea un topic de tipo “faceta” y la denominación (*topic name*) de esa misma faceta, 2) para cada uno de los términos de más alto rango, macrodescriptores o *top term*, dentro de cada faceta se crea un topic de tipo “término”, cuya denominación será la del propio término. A este topic se le asocia con la faceta a la que pertenece empleando el tipo de asociación “pertenece a la faceta”, 3) para cada uno de los términos pertenecientes al macrodescriptor se crea un *topic* de tipo

⁵²¹ AHMED, K. Topic maps for repositories. En *XML Europa 2000*. En: <http://www.gca.org/papers/xml europe2000/pdf/s29-04.pdf>. Consultado el 15/07/2009.

⁵²² DIJCK, Peter van. *Introduction to XFML*. En: <http://www.xml.com/pub/a/2003/01/22/xfml.html?page=2>. Consultado el 15/07/2009.

⁵²³ Id.

“término” cuya denominación (*topic name*) es la del propio término y relacionado con el topterm por medio de un tipo de asociación jerárquica “genérico/específico”.

4.4.3.1 Holística del modelo *Topic map* en el ámbito jurídico.

En las bases de datos jurídicas, y en general en la información digital, es habitual que los usuarios cuando emplean los motores de búsqueda obtengan mucho ruido informativo, páginas que poco o nada tienen que ver con sus demandas informativas; o bien puede que la inexistencia de enlaces hipermediales entre documentos multimedia de una misma materia impida la navegación entre los mismos.

Cuando se le pide a un navegador buscar por una ley fundamental como la Constitución española de 1875, esa ley es un *topic* de búsqueda sobre la cual el usuario espera poder localizar información de algún modo relacionada con su demanda. El navegador podría encontrar el texto completo de la Constitución, o bien una fotografía de la misma publicada, pero es muy probable que se pierda entre las miles de páginas entre las cuales “constitución de 1875” es únicamente una cadena de caracteres que se localiza en documentos donde únicamente se menciona el tema de forma superficial o incluso tangencial sin que se trate del tema central de la publicación.

De forma similar cuando se busca la materia “constitución de 1875” en el índice de un libro, se tiene la esperanza de localizar páginas cuyo descriptor sea éste, porque se presume la existencia de un vínculo entre lo que el usuario tiene en mente y los objetos informativos. Un conjunto de vínculos es una representación de la estructura de los objetos conceptuales del usuario para poder compartirlos con la red de ordenadores. Es lo que Fensel definió como ontología⁵²⁴:

La organización conceptual de la red de un modo explícitamente legible para una máquina.

Por tanto, las ontologías son estructuras conceptuales compartibles, escalables y reutilizables que inciden en la organización del conocimiento, especialmente en dominios bien estructurados como el Derecho. Es por ello que ya existen ontologías jurídicas que

⁵²⁴ CASANOVAS, P. *Ontología jurídicas profesionales. Sobre “conocer” y “representar” el derecho*. En: http://www.leibnizsociedad.org/secciones/mater/pon/textos/ontologias_pompeu.pdf. Consultado el 15/07/2009.

han permitido la construcción de prototipos de asistencia, búsqueda y organización documental de la información almacenada en las bases de datos jurídicas⁵²⁵.

Podemos situar las ontologías jurídicas dentro de las ontologías de dominio pero con dos diferencias⁵²⁶, por un lado se hace necesario contar con un lenguaje de especialidad que denominamos lenguaje jurídico; y por otro lado, se debe tener en cuenta en la construcción de la ontología el modelo particular del área del derecho a ser escogido. Por tanto, el proceso de diseño en el ámbito legal necesita de un nivel intermedio donde sea posible relacionar un conjunto de conceptos, implícita o explícitamente, con una colección de documentos jurídicos, estructurados de acuerdo con un modelo jurídico y cuyos contenidos son expresados con una sintaxis jurídica. Existe un nivel interpretativo, enlazado con las teorías generales del derecho, que es intermedio entre las ontologías de alto nivel y las específicas de dominio. También podemos encontrar implícitamente esta distinción entre un nivel de ontología y un nivel de aplicación en el modelado cognitivo, en el cual las categorías, *topics* y *occurrences* son diferenciados. Pero la característica más llamativa de las ontologías jurídicas construidas hasta ahora, es que la capa intermedia está explícitamente ocupada por construcciones conceptuales procedentes de las teorías generales del derecho, en lugar de los conceptos factuales de los dominios.

Hasta el momento, no se han construido muchas ontologías en el ámbito jurídico en el ámbito internacional. Un modo de describir la situación actual es el trabajo realizado por Valente⁵²⁷, quien por medio de la identificación de las principales ontologías jurídicas, como vemos en la tabla siguiente se han reconocido trece en los últimos años de investigación, ha aclarado algunos de los propósitos que intervinieron en el diseño de las ontologías jurídicas, entre los que destacan: la recuperación de información, la recuperación de estatutos, los hiperenlaces a normas, la gestión del conocimiento o el razonamiento legal. No obstante, el dominio jurídico es muy dependiente de las regulaciones jurídicas de cada territorio jurídico, razón por la cual algunas de las ontologías jurídicas base están orientadas a compartir un conjunto común de nociones legales. Estas ontologías se encuentran en un nivel de conocimiento general compartido en el ámbito del derecho, pero existe un tipo de

⁵²⁵ VALENTE, A. Types and roles of legal ontologies. En: Benjamins, V.[et al.]. *En Law and the Semantic Web. Legal ontologies, methodologies, legal information retrieval, and applications*. Berlin : Springer-Verlag, 2005. p.65ss.

⁵²⁶ CASANOVAS, P. [et al.] Semantic Web: a legal case study. En Davies, J. ; Studer, R. ; Warren, P. *Semantic Web Technologies*. Chichester : Wiley, 2006. p.262.

⁵²⁷ VALENTE, A. Op. cit., p.67.

conocimiento legal específico que pertenece a la cultura jurídica de cada país y que no se encuentra en este nivel intermedio de Ontología jurídica base.

Ontología	Aplicación	Tipo	Función	Carácter
Lenguaje de McCarty del discurso legal	Lenguaje general para la expresión de conocimiento legal.	Representación de conocimiento, altamente estructurado	Comprensión de un dominio	General
Ontología funcional de derecho de Valente y Breuker.	Arquitectura general para la resolución de problemas legales.	Base de conocimiento con la aplicación Ontolingua. Altamente estructurado.	Comprensión de un dominio, razonamiento y resolución de problemas.	General
Marco ontológico de Van Kralingen y Visser.	Lenguaje general para la expresión de conocimiento legal.	Representación de conocimiento moderadamente estructurado	Comprensión de un dominio.	General
Modelo de conocimiento jurídico basado en Mommer.	Lenguaje general para la expresión de conocimiento legal.	Base de conocimiento en inglés altamente estructurado.	Comprensión de un dominio.	General
Ontología base LRI de Breuker y Hoekstra.	Adquisición de conocimientos de apoyo para ontologías del dominio jurídico.	Base de conocimiento en OWL/RDF usando Protégé. Moderadamente estructurado.	Comprensión de un dominio.	General
Ontologías de conocimiento legal profesional de Benjamins y Casanovas.	Sistema FAQ inteligente para la recuperación de información para jueces.	Base de conocimiento en Protégé. Moderadamente estructurado.	Indización semántica y búsqueda	Dominio

Topic Maps y Tecnologías de la Información Jurídica: un modelo de recuperación de información para bibliotecas digitales semánticas jurídicas.

Ontologías de los Códigos franceses de Lame.	Recuperación de información legal	Base de conocimiento orientada al procesamiento del lenguaje natural. Ligeramente estructurada.	Indización semántica y búsqueda	Dominio
Ontología del fraude financiero de Leary, Vanderverghe y Zeleznikow.	Ontología para la representación de casos de fraude financiero.	Base de conocimiento en UML. Ligeramente estructurado.	Indización semántica y búsqueda	Dominio
JurWordNet de Gangemi, Sagri y Tiscornia.	Extensión de Wordnet al dominio legal.	Base de Conocimiento léxico. Ligeramente estructurado.	Organiza y estructura información	General
Ontología italiana del crimen de Asaro	Schema para la representación de los crímenes según el derecho italiano	Base de conocimiento en UML. Ligeramente estructurado.	Organiza y estructura información	Dominio
Ontología CLIME de Boer, Hoekstra y Winkel.	Sistema de asistencia legal para derecho marítimo.	Base de conocimiento en Protégé y RDF. Moderadamente estructurado.	Razonamiento y resolución de problemas.	Dominio
Ontología de causalidad legal de Lehman, Breuker y Browver.	Representación de la causalidad en el ámbito legal.	Base de conocimiento. Ligeramente estructurada.	Comprensión de un dominio.	Dominio
Ontología de derechos de la propiedad intelectual de Delgado.	Integración de Schemas que definen lenguajes para la expresión de derechos y diccionarios de	Base de conocimiento en sintaxis OWL.	Interoperabilidad entre los sistemas de gestión de derechos digitales.	Dominio

derechos.

Tabla 7. Tabla de las principales ontologías jurídicas publicada por Valente.

En España, no existen muchos ejemplos de aplicaciones de ontologías jurídicas; sin embargo, desde el campo de la informática jurídica se han desarrollado algunas experiencias en el ámbito judicial. Concretamente, en el seno de la Universidad Autónoma de Barcelona se encuentra el IDT (Institute of Law and Technology)⁵²⁸, que conjuntamente con la empresa iSOCO⁵²⁹ han desarrollado en el marco de un proyecto del Ministerio de Industria (PROFIT) una ontología de asistencia a las consultas de los nuevos jueces en el comienzo del ejercicio de sus cargos denominada IURISERVICE⁵³⁰.

Esta ontología jurídica, para la mejora de la organización, comunicación y transmisión de los contenidos de los documentos, no trata simplemente de gestionar la información, sino de gestionar el conocimiento, dado que proporciona la capacidad de discriminar entre los diversos significados contenidos en los mismos. Este objetivo pretende resolver el problema que se evidenció cuando en una encuesta realizada sobre los jueces recién titulados se identificaron los problemas más frecuentes a los que tienen que enfrentarse en sus primeras asignaciones, cuando por falta de experiencia, recurren a sus mentores para solicitar consejo profesional. Este sistema tradicional de compartir conocimiento judicial lleva a situaciones contradictorias para el mismo problema, disminuyendo la eficacia del sistema legal.

Para dar respuesta a esta situación se diseñó un sistema inteligente de preguntas frecuentes, FAQ's, que contiene miles de preguntas legales con sus correspondientes respuestas. Estas preguntas y respuestas contienen la experiencia judicial de más de 400 jueces de toda España. El sistema, Iuriservie, ofrece soporte a dos necesidades judiciales: 1) Un sistema de búsqueda de preguntas frecuentes, y 2) un sistema de búsqueda de casos legales. El sistema de búsqueda de preguntas frecuentes permite que el usuario navegue por la relación de

⁵²⁸ INSTITUTE OF LAW AND TECHNOLOGY. En: <http://idt.uab.cat/> Consultado el 15/07/2009.

⁵²⁹ iSOCO es una empresa surgida del instituto de inteligencia artificial del CSIC. En: http://www.isoco.com/innovacion_proyectos_nacionales.htm. Consultado el 15/07/2009.

⁵³⁰ BENJAMINS, R.; CASANOVAS, P.; CONTRERAS, J.; LÓPEZ COBO, J. M.; LEMUS, L. Iuriservice: An Intelligent Frequently Asked Questions System to Assist Newly Appointed Judges. En BENJAMINS, V.R., CASANOVAS, P. BREUKER, J., GANGEMI, A. (Eds.). *Law and the Semantic Web. Legal Ontologies. Methodologies, Legal Information Retrieval, and Applications*. Berlin : Springer Verlag, 2005. pp. 201-217.

preguntas-respuestas usando un interfaz de lenguaje natural, y la herramienta de gestión de casos legales ofrece una aplicación de búsqueda y navegación sobre bases de datos de jurisprudencia. En un escenario típico, un juez buscará primero una respuesta a una pregunta dada que formule en el sistema. La respuesta a esta pregunta, si no directa, da la información sobre qué buscar en la base de datos de jurisprudencia. Para que el sistema de preguntas frecuentes encuentre la pareja pregunta-respuesta que más se acerque a la pregunta realizada utiliza ontologías jurídicas. El primer paso es detectar la temática de la pregunta y clasificarla según los subdominios definidos en base a cálculos estadísticos. El siguiente paso es el proceso de búsqueda sobre ficheros, bases de datos, y otros repositorios, utilizando estrategias de búsqueda morfológicas, de sinónimos, búsquedas ortográficas y búsquedas semánticas. De esta forma se recuperan las preguntas según su significado legal, y no en base a las palabras utilizadas para formular la pregunta. El significado legal se define en la ontología del sistema.

Un sistema análogo fue el creado en el marco del proyecto europeo LOIS (Legal Ontologies for Knowledge Sharing), por el cual se diseñó una ontología jurídica a partir de la ontología léxica JurWordnet⁵³¹. La base del proyecto está en la creación de un índice interlingüístico (ILI) que relaciona los synsets o conceptos del Eurowordnet de cada una de las lenguas de la Unión Europea con un mismo synset en italiano, posteriormente contrastado en sus significados con el Corpus de legislación italiana y glosados por expertos del derecho a partir de manuales jurídicos.

Su arquitectura global, mostrada en la figura siguiente, explica cómo a partir de relaciones de equivalencia se va construyendo el índice ILI compuesto de identificadores. Estos se obtienen a partir de técnicas de alineamiento semiautomáticas orientadas a la constitución de una colección multilingüe de términos o synsets jurídicos con los cuales generar un único identificador contextualizado por un conjunto de corpora jurídicos en las distintas lenguas que forman parte de JurWordNet⁵³².

Una segunda etapa es su integración con una ontología externa como es el tesoro Eurovoc, que en la gráfica denominamos “external ontology”. El propósito es ampliar las

⁵³¹ TISCORNIA, D. JurWordNet, semantic web of legal technology. En: <http://www.ittig.cnr.it/Ricerca/UnitaEng.php?Id=11&T=4>. Consultado el 15/02/2010.

⁵³² TISCORNIA, D. The Lois Project: Lexical Ontologies for Legal Information Sharing. En: <http://www.e-p-a-com.dlib/9788883980466/art14.pdf>. Consultado el 15/02/2010.

relaciones jerárquicas del JurWordNet con las relaciones tesaurales. De esta manera se logra una representación más precisa y explícita de los términos, ya que cada descriptor se asigna a una clase ontológica con sus correspondientes relaciones propias de la clase⁵³³.

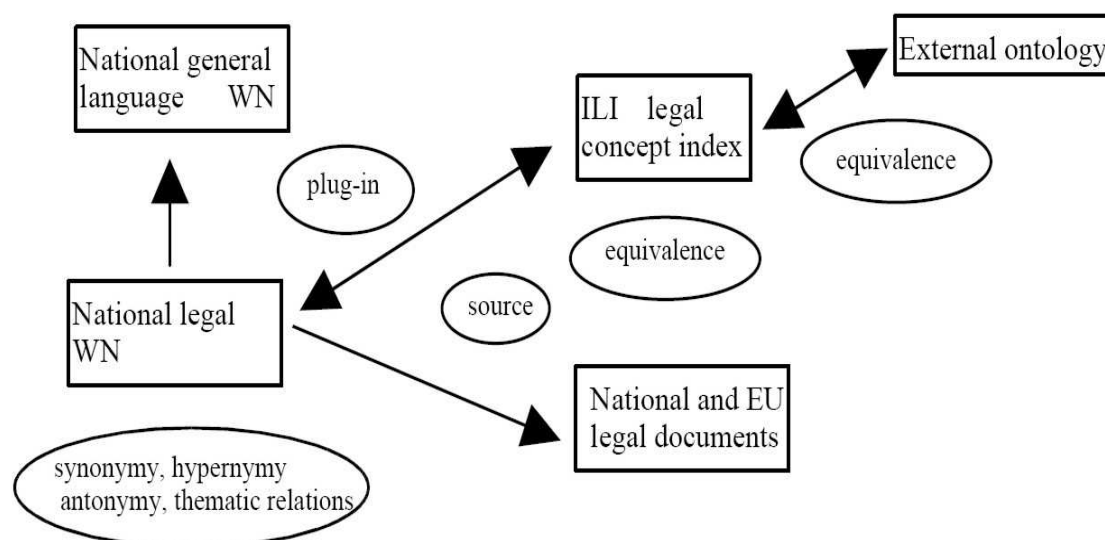


Figura 37. Diseño de la arquitectura de una ontología jurídica: Proyecto Lois.

En cuanto a la interfaz de búsqueda para este tipo de ontologías léxico-jurídicas, normalmente ofrecen una aplicación de recuperación en bases de datos de jurisprudencia y de decisiones legales a través de las ontologías. Estas aplicaciones permiten que los juristas realicen sus búsquedas en función de distintos filtros: por organización judicial, por temática, jurisdicción, fechas, número de sentencia, y algunos conceptos adicionales definidos en la ontología. Este subsistema es parecido a otras bases de datos jurídicas, pero mejorado por el uso de tecnologías semánticas que permiten expandir las preguntas realizadas en función del contexto.

⁵³³ Ibid., p.201.

Figura 38. iFAQ. Interfaz de búsqueda para JURISERVICE.

Entre los beneficios que aportan las ontologías jurídicas, según Casanovas, al ejercicio de la práctica jurídica están⁵³⁴:

- 1) Gestión pragmática de conocimiento profesional relevante para los jueces en su trabajo diario.
- 2) Promueve la consistencia en decisiones judiciales mediante la acumulación de conocimiento relevante.
- 3) Reducción del tiempo de reacción de nuevos jueces ante sus primeros casos.
- 4) Gestión del conocimiento (recuperación de preguntas y jurisprudencia) basado en la “comprensión” del contexto más que en palabras concretas.

Hasta el momento no existe en España ninguna experiencia de aplicación de topic maps ni en el ámbito jurídico ni en bibliotecas digitales, si bien existen ya numerosas experiencias de ello tanto en países europeos como en Estados Unidos, de lo que son buen ejemplo los congresos anuales que se celebran tanto en Oslo (*Topic maps*), como en Leipzig (TMRA), o en Asia (AToMS) sobre la materia⁵³⁵, incluidas experiencias de *topic maps* en el ámbito de la gobernanza⁵³⁶, la administración electrónica⁵³⁷, o las bibliotecas digitales; incluso tenemos

⁵³⁴ Ibid., p.217.

⁵³⁵ Se pueden acceder a las presentaciones de algunas de estas experiencias. En: <http://www.topic-maps.com/tmc/conferences.jsp> Consultado el 15/07/2009.

⁵³⁶ TVEIT, L. *How can local government use Topic Maps to make life easier for its citizens*. En: http://www.topic-maps.com/tmc/presentation.jsp?conf=TM2007&id=Lars_Tveit~TM2007~Bergen_Municipality. Consultado el 15/07/2009.

constancia de la presentación de algún trabajo acerca de la aplicación de *topic maps* en el ámbito jurídico⁵³⁸. Si bien no existen trabajos acerca de cómo diseñar *topic maps* en portales jurídicos, si se cuentan con numerosos ejemplos de ello en Noruega, destacando especialmente el portal del gobierno noruego⁵³⁹ que nos servirá como referencia para el diseño del portal del CEPC dirigido por un *topic map*.

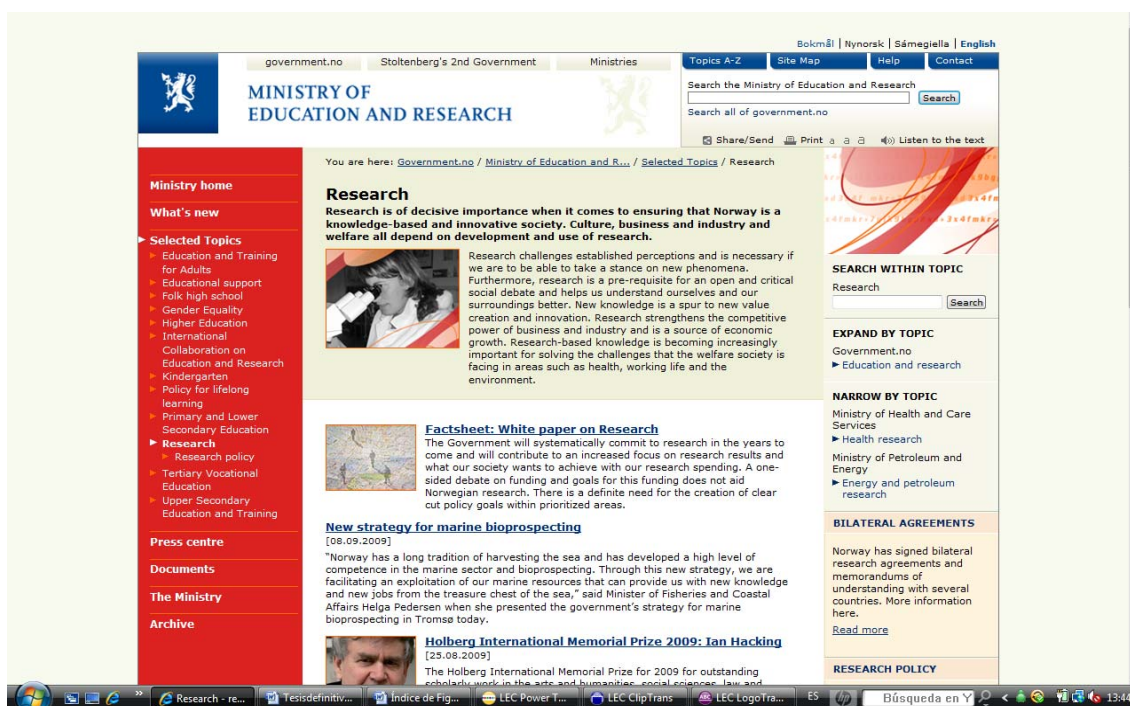


Figura 39. Portal organizado con topic maps del gobierno noruego.

Además, de cara al diseño de topic maps para el ámbito jurídico, contamos con las experiencias en el campo de las ontologías jurídicas, dado que un topic map es un tipo de ontología, y su diseño tiene muchas concomitancias⁵⁴⁰. De forma intuitiva podríamos expresar los topic maps en el ámbito jurídico de la siguiente manera. Supongamos que deseamos localizar todas las constituciones europeas aprobadas en España durante el siglo XIX con influencias del derecho alemán. No hay modo de formular una consulta como

⁵³⁷ AARTS, M. *Knowledge models for government information*. En: <http://www.topic-maps.com/tm2008/aarts.pdf>. Consultado el 15/07/2009.

⁵³⁸ BING, J. *Knowledge-based Systems for Representing Legal Norms*. En: http://www.topic-maps.com/tmc/presentation.jsp?conf=TM2009&id=Jon_Bing~TM2009~Knowledge-based_Systems_for_Representing_Legal_Norms. Consultado el 15/07/2009.

⁵³⁹ GOVERNMENT.NO. En: <http://www.regjeringen.no/en.html?id=4>. Consultado el 15/07/2009.

⁵⁴⁰ GARSHOL, L. *Modelling with Topic Maps*. En: http://www.topic-maps.com/tmc/presentation.jsp?conf=TM2007&id=Lars_Marius_Garshol~TM2007~Modelling_with_Topic_Maps. Consultado el 15/07/2009. Seminario al que asistió el autor de esta tesis.

ésta en un motor de búsqueda comercial como Google, ni en el motor de búsqueda de un OPAC. Se puede intentar emplear una colección de palabras-clave pertinentes como “Constitución+Alemania+España+siglo XIX”, pero el resultado es: No se ha encontrado ningún resultado para “Constitución Alemania España siglo XIX”. Además de no recuperar la información que se le solicita, me ofrece algunos resultados que poco tienen que ver con mi demanda informativa. Es más, aunque la búsqueda fuera muy precisa todavía tendría que seleccionar, de entre todos los documentos resultantes, los trabajos en los que el usuario podría localizar la respuesta exacta a su consulta. Sería mucho más sencillo si pudiéramos efectuar la consulta en un índice de todas las constituciones “escritas por” juristas de carrera asociado con formación en lengua alemana y con “estancias en Alemania”.

La diferencia clave entre ambas aproximaciones está en que la primera utiliza un índice a texto completo construido a partir de una colección de textos, mientras que la última emplea un índice que encapsula la estructura de un campo de conocimiento. Esta última solución es un topic map, una red estructurada de hiperenlaces aplicada a un repositorio de recursos de información. Cada uno de los nodos de la red representa un topic con una denominación. Los enlaces conectores de topics expresan las associations entre los mismos. Por tanto, los índices, en realidad, son formas sencillas de topic maps, al igual que los glosarios o los tesauros. Y todos estos lenguajes son ontologías en tanto que procuran representar un campo de la realidad.

El *topic map* jurídico modela un dominio del conocimiento jurídico por medio de los topics del ámbito jurídico, sus interrelaciones o associations y las occurrences de los recursos de información pertinentes. Junto a estas características se adjuntan otras que facilitan su extensión y gestión como los *topic types*, las *association types*, las *occurrences roles*, la identidad, las facetas y el *scope* o contexto.

En el contexto de un glosario jurídico, un topic representaría materias como “Gobernanza”, “subsidiariedad”, “déficit democrático”, cualquier topic susceptible de tener una entrada en el glosario con independencia de las características específicas que puedan tener, pero acerca de la cual pueda afirmarse algo. Sin embargo, aquello que uno elige considerar como topic, variará de acuerdo con las necesidades de la aplicación, la naturaleza de la información, y los usos a los cuales el topic map esté destinado.

Un claro ejemplo con el que se puede generar un topic map jurídico a partir de un glosario del mismo dominio lo constituye el glosario del portal Europa de la Unión Europea⁵⁴¹. Si buscamos el *topic* “gobernanza”, obtenemos una definición y los topics relacionados.



Figura 40. Glosario de Derecho Comunitario del Portal Europa.

Si diseñáramos, o más bien, trasladáramos las propiedades del glosario a un *topic map*, supondría tener un topic map de Derecho Comunitario con los siguientes grupos globales de características:

i) El primer grupo es el *name* de cada una de las entradas del glosario, las cuales pueden tener cero o más nombres, cada uno de los cuales válido en un contexto particular. La norma ISO 13250 designa el *topic name* como una forma conjunta de un *base name* y sus posibles variant names⁵⁴². Cada topic constará al menos de un *base name*, que es la forma básica del nombre y es siempre una cadena de caracteres. Cuando una aplicación elige emplear un topic name particular para designar un topic, el *base name* nos proporciona la cadena literal para la aplicación a emplear, a no ser que exista una denominación alternativa más

⁵⁴¹ Glosario del Portal Europa. En: http://europa.eu/scadplus/glossary/governance_es.htm. Consultado el 15/07/2009.

⁵⁴² ISO/IEC 13250 Topic Maps [second edition]. En: http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0322_files/iso13250-2nd-ed-v2.pdf. Consultado el 15/07/2009.

adecuada en el contexto semántico en que se aplique, el *variant name*. El *variant name* es la forma alternativa de un *basename*, pero optimizado para un propósito específico tal como la ordenación (*sortname*) o la visualización (*displayname*). Una aplicación *topic map* selecciona los *variant names* en función de la evaluación de sus parámetros contextuales; de forma que, los *base names* están sometidos a las restricciones de las denominaciones *topic* que prohíben el procesamiento de un *topic map* a partir de topics con el mismo *base name* en un mismo contexto.

Estos parámetros son información en forma de colección de topics que expresan el contexto apropiado de procesamiento para un *variant name*. Así si seleccionamos un *topicname* particular, una aplicación *topic map* puede examinar las variantes de sus nombres con el fin de seleccionar las formas más adecuadas para ese nombre o bien el *topic* en distintos idiomas. De hecho, nuestro glosario ya recoge las distintas lenguas en las que se puede recuperar por ejemplo el término “Gobernanza”.

ii) El segundo grupo de características es el *topic type*. Un topic puede estar subsumido en uno o muchos *topic types*. Así por ejemplo, Gobernanza es un topic del tipo “Sociedad de la Información” y también un topic del tipo “Justicia, Libertad y Seguridad”. Es decir, los *topic types* expresan relaciones cuya tipología se expresa por una relación clase-instancia o relación del tipo “es una”. Los *topic types* son en sí mismos topics, de forma que para que puedan ser empleados para la tipificación se hace necesario declararlos como topics en el *topic map* en que se inserten. Característica que nos permite ampliar la capacidad semántica del *topic map* en sí mismo.

iii) El tercer grupo de características es el correspondiente al *topic occurrence* y *occurrence role*. Un *topic* puede tener una o más *occurrences*. La *occurrence* de un topic es un hipervínculo a uno o más recursos de información que de alguna manera son relevantes para la materia que el topic representa⁵⁴³. Así por ejemplo podríamos encontrarnos con un artículo acerca de la gobernanza, la fotografía de una edición de un texto sobre gobernanza, o un vídeo descriptivo del desarrollo de la gobernanza desde el nacimiento de la Comunidad Económica Europea hasta la actual Unión Europea, o sencillamente podría tratarse de una mención de la misma en el contexto de otro concepto como pueda ser el contexto de la administración electrónica en España, o bien puede tratarse de una presentación en otros

⁵⁴³ Ibid., p.6.

recursos de información para los cuales tiene relevancia la gobernanza, como un powerpoint sobre el papel de las tecnologías de la información en el funcionamiento del Parlamento Europeo. Normalmente, todos estos recursos se encuentran fuera del documento topic map, pero están señalizados por medio de los muchos mecanismos de direccionamiento que hay en internet como Hytime o XPointer. Las occurrences pueden ser agrupadas en distintas categorías o clases para representar distintas formas documentales o el soporte material en el que se encuentren los recursos: libros, ilustraciones, comentarios, papel, digital. A estas clases se las conoce por *occurrence role*⁵⁴⁴, que como los *topic types*, son considerados topics, aunque las occurrences no lo sean.

Los *topics* y las *occurrences*, conjunta y articuladamente, son parte de un modelo de representación de las materias de una colección orientado a la localización guiada de recursos a través de la navegación entre una colección de documentos heterogéneos y dispersos. Al igual que los índices, los términos de sus entradas son los topics y los números de página son sus occurrences. Sin embargo, cuenta con un conjunto de características adicionales, como el concepto de *occurrence role*, que introduce la posibilidad de diferenciar entre los distintos tipos de referencias de cada una de las *occurrences*.

En el caso de un *topic map*, construido a partir de un glosario con enlaces a los documentos (*occurrences*) de la síntesis de legislación de la Unión Europea, el empleo de distintos tipos de letra para las occurrences indica distintos roles en páginas diferentes. Así, si tomamos entre los términos del margen izquierdo de nuestro glosario el concepto de “Sociedad de la Información” nos dirigimos a una descripción de los objetivos de la Unión Europea respecto a la materia, pero si tomamos el término dentro del propio glosario se nos proporciona una explicación de su significado. Es decir, “Sociedad de la Información” se muestra dos veces con tipografías diferentes en nuestro glosario y, por tanto, con informaciones distintas pero complementarias.

En definitiva, se concibe un glosario, y por tanto cualquier tipo de índice, como una forma sencilla de *topic map*. Pero, ¿qué ocurre cuando deseamos trabajar simultáneamente con el glosario de derecho del portal Europa, el índice alfabético de su síntesis legislativa y el mapa de sitio?, ¿tendríamos que construir tres topic maps y realizar hiperenlaces entre los términos de cada uno de estos tres índices?

⁵⁴⁴ Ibid., p.3.

Nuestra experiencia como documentalistas nos permite saber que la manera de facilitar a nuestros usuarios el acceso a los documentos de una colección es poniendo a su disposición un conjunto de índices⁵⁴⁵: de nombres, de lugares, de materias, de figuras, de fotografías, de gráficas, etc.; que dotan al usuario de capacidad para navegar, en el sentido de capacidad para localizar información por sus atributos y relaciones⁵⁴⁶. Los índices permiten enlazar una instancia, con sus atributos y relaciones, con las páginas en las que se localiza. Sin embargo, ¿qué ocurriría si fuéramos capaces de relacionar cada una de las instancias de un índice con las instancias de los otros índices? El principal problema con el que nos encontraríamos sería la falta de normalización de los índices, lo cual repercute en su interoperabilidad. Pero con los topic maps es posible fusionar índices, con independencia de la finalidad con la que hayan sido construídos, obteniendo un metaíndice que facilita al usuario no sólo una navegación explícita entre más instancias, sino que sobre todo proporciona una navegación contextual derivada de la constitución de relaciones semánticas entre las instancias⁵⁴⁷.

Un claro ejemplo de ello lo tenemos en el topic map que hemos diseñado para el Centro de Estudios Políticos y Constitucionales. En el margen izquierdo se muestra el índice maestro con el cual se integran todos los índices: i) índice de *topic types*, ii) índice de temas, etc.

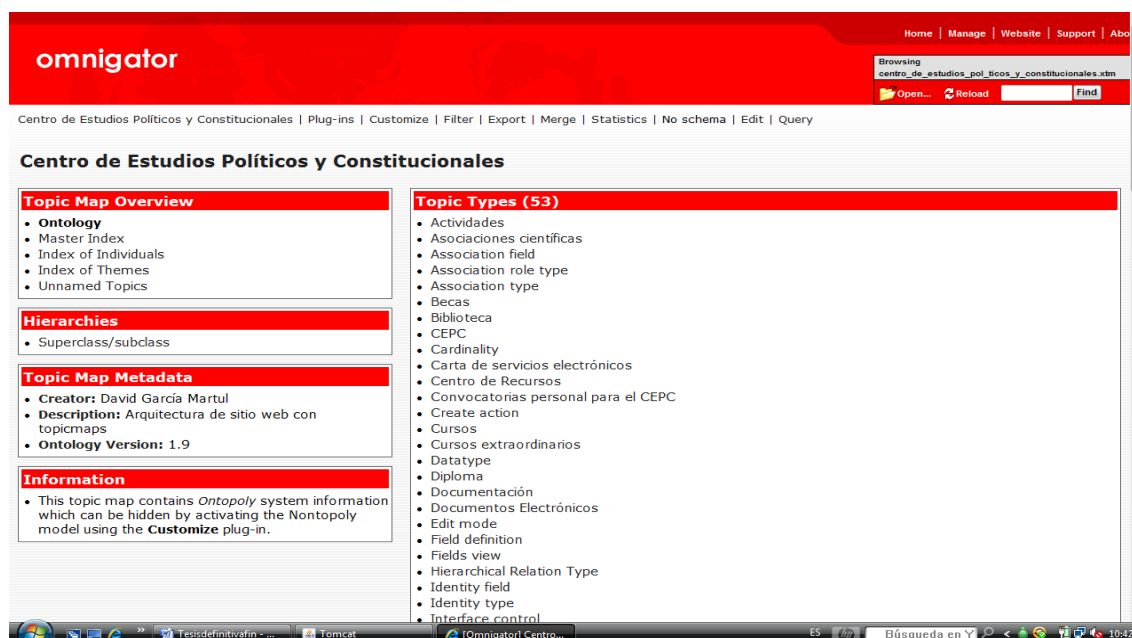


Figura 41. Topic Map para el Centro de Estudios Políticos y Constitucionales. Elaboración propia.

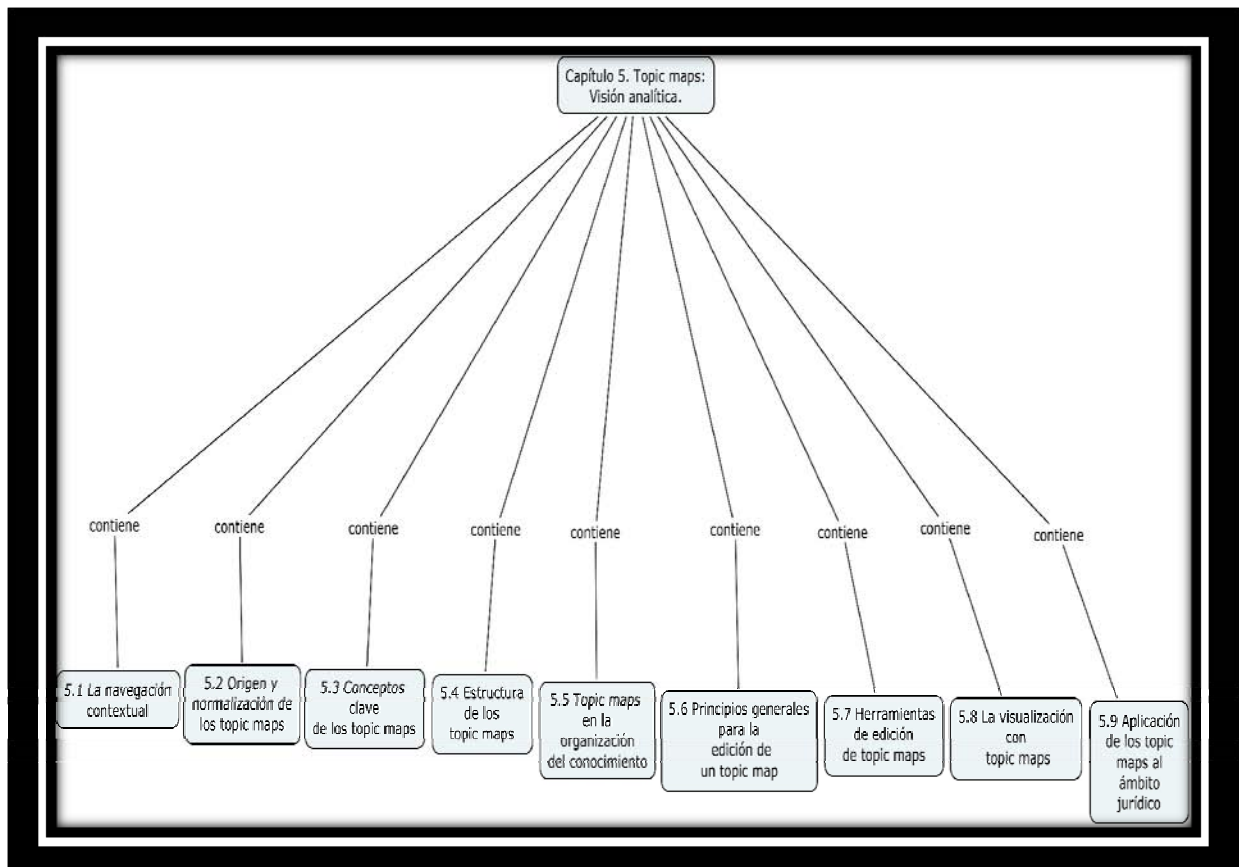
⁵⁴⁵ TAYLOR, A. *The organization of information*. Op.cit., p.40.

⁵⁴⁶ PASSIN, T. *Explorer's guide to the semantic web*. Greenwich : Manning, 2004, p.62.

⁵⁴⁷ Ibid, p. 63.

CAPÍTULO 5.

Topic maps: Visión Analítica



Introducción.

Un espacio de información, especialmente si es tan extenso como la World Wide Web, necesita de un modelo de representación capaz de preservar el contexto en la navegación y recuperación de los objetos de información. Para ello, si fuera posible indicar la materia de cada objeto de información de la web, además de la semántica que se establece en su relación con otros objetos relacionados, entonces la recuperación no sólo sería más fácil sino que sería mucho más precisa. Esta es la idea que hay detrás de la web semántica como la concibió Berners-Lee y el consorcio W3C.

Topic maps es una tecnología normalizada para la descripción de estructuras de conocimiento y mejora de la localización de la información⁵⁴⁸. Se basa en un modelo formal que subsume los modelos tradicionales para la localización de información, tales como índices o tesauros, al mismo tiempo que proporcionan un conjunto de capacidades adecuadas a las características de la documentación digital. Capacidades que intentan resolver la dificultad de navegar por Internet causada por la heterogeneidad de documentos por los que se navega; pero sobre todo, por la desconexión semántica que existe entre ellos. Este problema lo afrontan proporcionando:

- I) un modelo de datos flexible para la recogida de conocimiento y,
- II) una sintaxis normalizada que hace accesible el conocimiento a la comunidad de usuarios en concurrencia con aplicaciones capaces de compilar esta sintaxis.

Por tanto, son una norma para la descripción de estructuras de conocimiento, representadas en forma de *topics*, relaciones entre los *topics*, y relaciones de estos con los documentos. Posibilitan la navegación entre los documentos sobre la base de las materias y las relaciones entre las mismas. Su concepción se desarrolló para facilitar la navegación y gestión en repositorios de información masivos y heterogéneos, a través de una capa de navegación construida con independencia de la forma y formato de los documentos propiamente dichos⁵⁴⁹.

Este diseño proporciona un gran poder al modelo, permitiendo que entre otras cosas el *topic map* pueda ser autodocumentado en el sentido de que sus materias cuentan con propiedades que las identifican⁵⁵⁰; es decir, dos sistemas de información pueden referirse a las propiedades de una materia de distinta manera. Por ejemplo: la materia “constitucionalista” en un sistema de información puede aparecer en otro sistema como “especialista en leyes fundamentales”, ambas materias son la misma a pesar de que cada una de ellas cuente con propiedades que las identifican, pero son sus propiedades las que permiten decidir si ambas materias son la misma.

⁵⁴⁸ PEPPER, S. Topic Maps. En BATES, M ; MACK, M. *Encyclopedia of Library and Information Sciences*. Los Angeles : CRC press, 2009. En: http://www.ontopedia.net/pepper/papers/ELIS-Topic_maps.pdf. Consultado el 15/12/2009.

⁵⁴⁹ AHMED, K. [et al.]. *Professional XML Metadata*. Chicago : Wrox Press, 2001. pp.269-270.

⁵⁵⁰ DURUSAU, P. *Understanding Subjects and Subject Proxies*. En: http://www.durusau.net/publications/Understanding_Subjects_and_Subject_Proxies.pdf. Consultado 15/12/2009.

En la ontología de un *topic map*, el conjunto de conceptos que lo constituyen, se define en términos de *topics* en un mapa de conocimiento. El mapa representa a la propia ontología, proporcionándole más funcionalidad y flexibilidad cuando es empleado para la navegación o la consulta.

Los *topic maps* comenzaron como un modo de representación de las estructuras de conocimiento reproducidas en los índices tradicionales de los libros con el fin de resolver los problemas de gestión de información implicados en la creación, mantenimiento y procesamiento de índices para documentación muy compleja. Mientras el modelo iba siendo desarrollado, su ámbito iba ampliándose hasta abarcar otros tipos de ayudas navegacionales, tales como glosarios, tesauros y referencias cruzadas⁵⁵¹.

Uno de los aspectos más innovadores de los *topic maps* fue el empleo de hiperenlaces y mecanismos de direccionamiento independientes. Esto permite separar los documentos de los índices que permiten su recuperación, de forma que puede dejar de hacerse necesario escribir las direcciones de acceso a los documentos electrónicos.

No obstante, en vez de limitarse a reproducir las propiedades de los índices impresos, el modelo *topic map* permite generalizarlos, extendiéndolos a todo tipo de documentos por su contenido, y por tanto permitiendo una navegación mucho más ágil y extensa. Con ellos un usuario puede recorrer el espacio multidimensional de las materias incluidas en un dominio de conocimiento, antes de decidir qué recursos de información son pertinentes a su consulta. Esto evita tener que revisar grandes repositorios de información para resolver sus demandas informativas. Análogamente, las consultas son mucho más precisas que las simples búsquedas a texto completo. De ser un útil pero infrautilizado recurso para la búsqueda a texto completo, los índices, organizados con *topic maps*, pasan a convertirse en un mapa navegable para la difusión y consumo de información.

Las posibilidades que ofrecen para la recuperación semántica de información van más allá de las tradicionalmente asociadas a los índices. Su estrecha relación con las redes semánticas proporciona la idea de que aún sin *occurrences* asociadas, pueden llegar a constituirse en un valioso sistema de organización del conocimiento.

⁵⁵¹ DURUSAU, P. *Babel and Topic Maps*. En: http://www.durusau.net/publications/Babel_and_TopicMaps.pdf, p.2. Consultado el 20/02/2010.

Su capacidad para la codificación de estructuras de conocimiento, así como para relacionarlas con los recursos informativos u occurrences, evidencia un papel destacado para los *topic maps* en la esfera de la gestión del conocimiento. Así, los *topic maps* pueden ser empleados para la representación de la interrelación de funciones, productos y procedimientos que constituyen la memoria corporativa, para su posterior interconexión con la documentación correspondiente⁵⁵².

5.1. La navegación contextual.

Los *topic maps* son un esquema de modelado de relaciones que pueden ser establecidas entre conceptos o *topics* representativos de la realidad. Un *topic* puede ser de hecho la representación de cualquier concepción de la realidad que nos formulemos. Sin embargo, esta representación de los conceptos carece de significado en tanto no seamos capaces de insertarla en un marco contextual concreto. El contexto es algo que se puede definir de forma casi intuitiva, pero los ordenadores no entienden el contexto de empleo de los *topics* si no se les enseña a reconocerlos. Es por ello, que el diseño de bases de datos a través de modelos de representación como el modelo de Chen, o bien el diseño de ontologías se ha convertido en un campo en auge de la disciplina de la computación y mas concretamente de la inteligencia artificial, la recuperación de información y la organización del conocimiento.

Teniendo en cuenta esto último, podemos definir el objetivo primario de un *topic map* como el propósito de proporcionar un entorno contextual rico en el cual es posible navegar por un mapa de *topics* representativos de una colección de recursos de información⁵⁵³. Para los documentalistas, son en primer lugar una herramienta de navegación.

Tradicionalmente, la navegación ha sido sobre todo un proceso de deambular a través del hipertexto con escasas indicaciones que ayuden al usuario a fijar un trayecto que le conduzcan a satisfacer sus demandas informativas. Como mucho se le ofrece un directorio o una jerarquía de materias normalizadas que constriñen la capacidad del usuario para construir sus propias estrategias de búsqueda. Es por ello que se dice que la navegación a

⁵⁵² SIGEL, A. Topic maps in Knowledge Organization. En Park, J. ; Hunting, S. *XML Topic maps*. Boston : Addison-Wesley, 2002. pp.383-384.

⁵⁵³ BROWN, P. *Information Architecture with XML: a management strategy*. London : John Wiley, 2003. pp.290-291.

través de mapas de sitio es más bien una exploración guiada que un proceso de autoaprendizaje⁵⁵⁴.

Con los *topic maps*, el potencial de los hiperenlaces se combina con una estructura sistemáticamente articulada de *topics* navegables y un interfaz de usuario flexible capaz de facilitar la usabilidad y accesibilidad por medio de un entorno contextual necesario para efectuar la navegación semántica⁵⁵⁵.

La navegación semántica significa que al usuario se le indica en todo momento el contexto en el que se encuentra mediante la indicación de un contexto constituido por las relaciones entre *topics*, así como la designación que se hace de cada uno de ellos. De esta manera, se le proporciona al usuario un entorno más rico para decidir el sentido de su navegación.

Desde el punto de vista del usuario el éxito en la navegación depende de los referentes con que cuente para superar el vacío semántico⁵⁵⁶ entre el suceso informativo en el que se encuentra y aquel en el que desearía estar.

La idea de un suceso informativo nace de la teoría de la mediamorfosis⁵⁵⁷, según la cual el medio es el mensaje, o más bien el mensaje está en el medio y el espacio intermedio de transmisión entre dos nodos es vacío. De forma que la misión del documentalista pasa por tener las habilidades y destrezas para diseñar arquitecturas de información dinámicas, capaces de organizar la información en el momento de su recepción, recogiendo así la idea de la organización del conocimiento en entornos de información dinámicos.

⁵⁵⁴ Id.

⁵⁵⁵ RATH, H. *Semantic Resource Exploration with Topic maps*.

En: <http://www.uni-giessen.de/germanistik/ascl/gldv2001/proceedings/pdf/GLDV2001-rath.pdf>. Consultado el 15/12/2009.

⁵⁵⁶ Entendemos como vacío semántico lo que en la literatura científica se designa como “semantic gap”. Hemos decidido designarlo así haciendo una analogía con la propagación de la luz en el vacío. Si la luz contiene información, ésta sólo se contiene en el momento de su transmisión y no en un espacio físico concreto. Así el usuario obtiene la información en un “event” o suceso informativo porque la información ha dejado de ser permanente o lentamente cambiante para pasar a ser relativa al tiempo. Así el vacío semántico es ese espacio lleno de datos y vacío de información, en la que es el usuario quien organiza y contextualiza los datos en un momento o suceso particular para lograr la información que desea.

⁵⁵⁷ FIDLER, R. *Mediamorfosis: comprender los nuevos medios*. Buenos Aires : Granica, 1998.

Un precedente de este esquema lo encontramos en la teoría de la información de Shannon⁵⁵⁸, más concretamente con su teoría de codificación de la fuente. Empleando una descripción estadística para los datos, la teoría de la información cuantifica el número de bits necesarios para describir los datos, lo que se denomina la entropía de información de la fuente. Esta teoría es importante para tener en cuenta que durante la transmisión de la información se pueden perder datos a causa del ruido y en la codificación y decodificación de los datos. Lo cual es un reflejo del valor que el documentalista debe asignar en el diseño de una arquitectura de información a los distintos elementos y procesos que intervienen en la transmisión de la información.

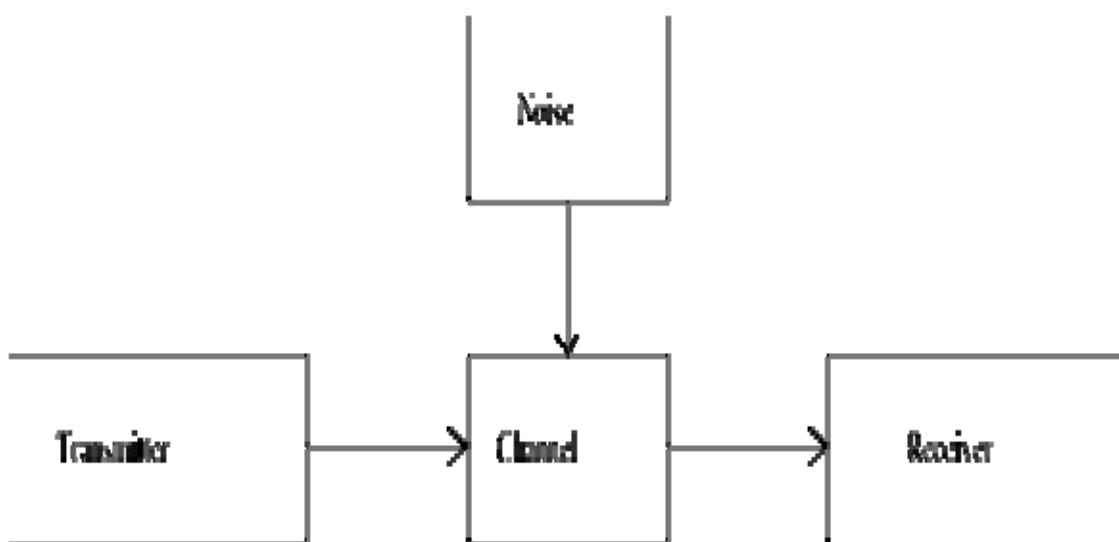


Figura. 42. Esquema del modelo matemático de información de Shannon.

Otro modelo teórico⁵⁵⁹ es la teoría de la información semiótica, como parte de la teoría del signo formulada por Peirce. Esta teoría, utilizada en lingüística computacional e inteligencia artificial, es un modelo que explica cómo es posible agrupar un conjunto de resultados bajo una colección de propiedades verídicas acerca de un objeto, de esta forma los signos podrían ser entendidos como designaciones (términos) o nociones (conceptos) de un esquema ontológico o taxonómico; es decir, como predicados y predicciones de los resultados. Si los resultados son capaces de agrupar objetos necesarios para alcanzar un fin,

⁵⁵⁸ SHANNON, C. A mathematical theory of communication. *Bell system technical journal*, july 1948. En: <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>. Consultado el 09/01/2010.

⁵⁵⁹ Peirce's Theory of Signs. En *Stanford Encyclopedia of Philosophy*. En: <http://plato.stanford.edu/entries/peirce-semiotics/>. Consultado el 29/10/2009.

entonces los signos podrían ser entendidos como fragmentos de órdenes indicativas de lo que hacer en una situación relativa a nuestro marco de objetivos.

En cuanto al proceso de navegación en un *topic map*, debemos señalar que se organiza con un esquema en tres niveles.

- 1) El inferior, corresponde a los recursos de información. Enlazando estos con sus correspondientes *topics*, se proporciona acceso directo al usuario desde sus demandas informativas a los recursos con independencia de donde éstos se encuentren, de forma que el *topic map* realiza una doble función de índice compuesto y de pasarela o gateway.
- 2) El segundo nivel corresponde al superior de los *topics*. El *topic map* no tiene por qué relacionarse con recursos de información, de forma que, la propia estructura organizativa de los *topics* puede servir como esquema básico para el diseño de otros esquemas de organización del conocimiento, tales como: taxonomías, tesauros o clasificaciones, a partir de la normalización y desambiguación de los términos y de las relaciones entre los mismos.
- 3) El tercer nivel del paradigma de los *topic maps* tiene que ver con la identidad de las materias. El concepto de *topic* se emplea de manera indistinta en campos como la inteligencia artificial o la computación, mientras que en la disciplina de la Documentación se opta por referirse a las materias; y a los términos en lingüística y terminología. Sin embargo, en análisis de contenido y lenguajes documentales existe el problema de que no siempre un *topic* o descriptor designan una materia de manera inequívoca. El concepto de PSI en *topic maps*, identificador público de materia, proporciona un instrumento para consensuar públicamente definiciones precisas de las materias de un dominio, y para posibilitar la relación de los *topics* con las materias⁵⁶⁰.

⁵⁶⁰ PEPPER, S. ; SCHWAB, S. *Curing the web's identity crisis*. En: http://www.ontopia.net/topic_maps/materials/identitycrisis.html. Consultado el 15/12/2009.

5.2. Origen y normalización de los *Topic maps*.

El desarrollo de los *topic maps* comenzó en 1991 como una iniciativa del grupo de Davenport, consorcio de vendedores de sistemas UNIX y editoriales como O'Reilly, para el desarrollo de un marco de gestión digital de manuales de software⁵⁶¹.

Los proveedores se encontraban bajo la presión de sus clientes de mejorar la coherencia en los índices de la documentación impresa. Existían inconsistencias sobre el empleo de términos en documentación y libros sobre sistemas informáticos UNIX, lo cual impedía incluir, por ejemplo, los documentos creados por la editorial O'Reilly sobre X-Windows en sus manuales de sistemas. El mayor problema residía en la manera de proporcionar índices maestros que pudieran ser mantenidos de forma independiente y fácilmente actualizables en una documentación técnica con contenidos muy cambiantes.

Para afrontarlo, el grupo de Davenport, se planteó la necesidad de dotar a los grandes grupos industriales de índices unificados para todos los manuales de programas de software que estaban adquiriendo a medida que iban automatizando los procesos industriales. Para ello, se hacía necesario un metaíndice que solo es posible en un mismo dominio de conocimiento para que éste tenga una coherencia mínima entre los índices que integra. Sin embargo, cada dominio de conocimiento tiene unos modelos implícitos de organización de la información, con lo cual se hace necesario no sólo unificar los *topics* que constituyen cada uno de los índices sino que también se hace necesario recoger el modelo que integra las materias de las fuentes informativas que indiza. El hecho de recoger estos modelos es lo que permite el proceso de unificación de índices.

El reto de crear índices maestros cambiantes hizo que en 1993 se crease un grupo llamado CapH⁵⁶² para llegar a un convenio sobre la aplicación de un lenguaje de marcado llamado HyTime para el empleo de hipertexto. Del lenguaje Hytime se publicó su norma en 1997 como ISO 10744 con el objeto de proporcionar a SGML propiedades de multimedia e hiperenlace, pero ya en 1991, uno de los padres de los *topic maps*, Newcomb, observó sus

⁵⁶¹ NEWCOMB, S. A perspective on the quest for global knowledge interchange. En Park, J. ; Hunting, S. *XML Topic maps. Creating and using topic maps for the web*. Boston : Addison-Wesley, 2002. pp.37-40.

⁵⁶² HyTime Users Group Home Page. En: <http://www.hytime.org/index.html>. Consultado el 15/12/2009.

propiedades para la transmisión de información relativa a la música en muy diferentes formatos a través del lenguaje SMDL que era una aplicación de HyTime⁵⁶³.

Una vez se constataron las posibilidades tan amplias ofrecidas por la navegación a través de hiperenlaces, el grupo CapH elaboró el modelo SOFABED como un *topic map*. Una vez que este modelo maduró lo suficiente como para ser aceptado por la ISO, se aprobó un borrador como base para una nueva norma internacional, que sería terminada en 1999 y publicada en enero del 2000 con el nombre de ISO/IEC 13250:2000 *Topic maps*. Una norma ISO que definía una sintaxis para *topic maps*. Esta sintaxis era una DTD SGML, que usaba la norma ISO 10744 HyTime para enlaces y asignación de direcciones, y por ello la sintaxis es conocida como *Hytopic maps* (abreviatura de HyTime *topic maps*). Cuando *Hytopic maps* fue completado, se plantearon tres cuestiones sobre la sintaxis:

No es una sintaxis XML. *Hytopic maps* fue especificada usando SGML, y como mucha gente emplea XML, se hizo necesario reconocer una versión XML.

No es una DTD completa. La sintaxis *Hytopic maps* no especifica cómo las referencias a documentos externos pueden ser representadas, ni qué sintaxis puede ser empleada para referencias internas. El resultado ha sido que cada programador de software *topic maps* ha realizado su propia versión *Hytopic maps*, derivada de la norma *Hytopic maps* DTD.

No utiliza URIs para referencias externas. Una sintaxis que no usa URIs no se integra bien en un contexto web, y se ha reconocido generalmente que era un importante requisito que los *topic maps* fueran capaces de hacerlo.

Con la necesidad de resolver estos problemas, añadido a la adaptación de los *topic maps* a la web, la organización *topic maps.org* decidió crear una nueva sintaxis *topic map* basada en XML y URIs. La sintaxis *topic map* creada se le denominó XTM (XML *topic maps*), y resolvía los problemas anteriormente enumerados. En octubre de 2001 la DTD de XTM fue aceptada en la ISO 13250, y por ello la segunda edición de ISO 13250 contiene ahora dos sintaxis: Hytime y XTM.

⁵⁶³ NEWCOMB, S. Standards. Standard Music Description Language Complies with Hypermedia Standard. *IEEE Computer* 24/7 (July 1991). pp.76-79.

Para llegar a la aprobación de esta norma se hizo necesaria la constitución de varios subgrupos, de los cuales uno de ellos asumió la tarea de definir una DTD en sintaxis SGML para el marcado del contenido de los manuales. Como resultado de esta labor se editó la DTD DocBook⁵⁶⁴. Otro subgrupo tuvo por tarea el desarrollo de un índice maestro digital para manuales de vendedores del sistema operativo X-Windows, para máquinas UNIX, y para un libro sobre este sistema operativo editado en O'Reilly. Este índice maestro se convertiría en una fuente imparcial y coherente para todos los índices específicos de cada una de las casas editoriales, ya que se creó a partir de la unión de todos estos índices. El subgrupo responsable pasó a denominarse SOFABED, que significa: Arquitectura Formal Abierta Normativa para Documentos Electrónicos Navegables. A pesar de parecer una tarea relativamente sencilla, tardaron nueve años hasta que se publicó en el año 2000 la norma ISO/IEC 13250, normativa de los *topic maps*, en donde se manifiestan los conceptos originalmente recogidos en el modelo del SOFABED⁵⁶⁵.

En estos nueve años el modelo del subgrupo SOFABED pasó a denominarse: Convenciones para la aplicación de HyTime (CApH), reconociendo así que la complejidad de describir un índice en formato electrónico se vería facilitada por la norma ISO/IEC 10744 HyTime a causa del potencial hipertextual que proporciona. El grupo CApH constató que las posibilidades del modelado de índices iban más allá de lo previsto en el modelo SOFABED y lo red denominó *Topic maps*. En 1995 la organización ISO aceptó el modelo como nuevo elemento de trabajo y en diciembre de 1999 el comité JTC1 SC34 de la organización lo aprobó como norma internacional ISO/IEC 13250.

En esta norma se especifican los conceptos *topic map* y se define el formato que permite a las distintas aplicaciones intercambiar la información de los *topic maps* de forma normalizada. Sin embargo, la norma del año 2000 se escribió con sintaxis HyTime, la cual no era muy apropiada para la web. Entorno ideal para cualquier norma de tecnologías de la información digital orientada a su organización⁵⁶⁶.

Así pues, la actual norma ISO 13250 define dos sintaxis de intercambio, XTM y Hytime, pero no explica cómo se relacionan entre sí. Existen numerosas diferencias, nada triviales,

⁵⁶⁴ DocBook. En: www.docbook.org. Consultado el 15/12/2009.

⁵⁶⁵ RATH, H. *The Topic maps Handbook*. Germany : Empolis, 2003. p.45.

⁵⁶⁶ Ibid., p.7.

entre las sintaxis que plantean problemas importantes de interoperabilidad⁵⁶⁷. Por ejemplo, la estructura de los nombres *topic* es diferente en las dos sintaxis. En Hytime la estructura de los nombres es como sigue⁵⁶⁸:

```
<!element (basename | sortname) -- Base name or Name to be used as sort key --
  - O (#PCDATA) -- String to be used as name/sorting sequence -- >
<!element dispname -- Displayable form of name --
  - O (#PCDATA|TMBrid)* -- String (or notation data embedded within a TMBrid
    conformant element) to be displayed as name -- >
<!attlist ( basename | sortname | dispname)
  scope -- Scope --
    -- References to a set of themes (topic
    links) to be added to the scope of the
    name characteristic specified in the
    content. --
  CDATA -- Reference --
    -- Reftype: topic+ --

  #IMPLIED -- Default: No themes are added via this attribute. -- >
```

En XTM la estructura es como sigue⁵⁶⁹:

```
<baseName>
  <scope> ... </scope>
  <baseNameString>... </baseNameString>
  <variant>
    <parameters>... </parameters>
    <variantName>... </variantName>
  </variant>
</baseName>
```

El problema es cómo relacionar los nombres de visualización y ordenación, además de los diferentes modos para designar sus correspondientes ámbitos. Esto es sólo un ejemplo de las diferencias entre ambas sintaxis, lo que dificulta su labor de mapeo.

Otro problema es que ambas sintaxis, en la norma ISO 13250, fallan en cuanto a la especificación de qué implementaciones se tienen que hacer en cada situación, ya que existen algunos puntos en donde éstas no están claras. En algunos casos los programadores han interpretado el texto de la especificación de manera diferente, y esto supone problemas de interoperabilidad. Si distintas implementaciones interpretan el mismo *topic* map de forma

⁵⁶⁷ BIEZUNSKI, M. ; NEWCOMB, S. ; BRYAN, M. *Guide to the topic map standardization*. En: <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0323.htm>. Consultado el 15/12/2009.

⁵⁶⁸ NEWCOMB, S. ; BIEZUNSKI, M. ; BRYAN, M. *The HyTime Topic maps (HyTM) syntax 1.0*. En: <http://www.jtc1sc34.org/repository/0391.htm>. Consultado el 15/12/2009.

⁵⁶⁹ PEPPER, S. ; MOORE, G. *XML Topic maps (XTM) 1.0*. En: http://www.topic_maps.org/xtm/1.0/index.html. Consultado el 15/12/2009.

diferente las aplicaciones *topic maps* sólo pueden trabajar con una única implementación, lo cual invalida el hecho de contar con una norma⁵⁷⁰.

Es por ello que el comité ISO SC34 revisó la norma ISO 13250 como una norma multiparte que resolviera los problemas descritos. Una parte clave de esta nueva edición de la norma era la que se conoce como *Standard Application Model* (SAM), un modelo de datos formal para *topic maps*. Este está basado en el mismo formalismo que el XML *information set*. Define la estructura permitida en *topic maps*, así como sus operaciones clave, tales como la unión y filtrado de duplicados. El SAM es lo que permite al SC34 resolver los problemas con las interpretaciones de las especificaciones, relaciona Hytime y XTM entre sí, y crea un fundamento tanto para el lenguaje de consulta de *topic maps* como para su lenguaje de restricciones o gramatical⁵⁷¹.

Además, el problema con la interpretación de las especificaciones de la norma ISO 13250:2000 y XTM 1.0 también fueron resueltas con la definición de nuevas especificaciones para las sintaxis Hytime y XTM basadas en el SAM. Las nuevas versiones de las especificaciones sintácticas describen cómo construir un caso del modelo SAM desde un documento con una sintaxis dada, pero no cambia su sintaxis. Esto es, utilizan declaraciones como “para cada *<topic>* en el documento, crea un ítem *topic*”, “para cada *<basename>* hijo del elemento *<topic>*, crea un ítem *basename* y lo añade a la propiedad *base names* del correspondiente ítem *topic*,” y así otros.

Reescribiendo las especificaciones sintácticas se resuelve el problema de la relación entre la sintaxis XTM y Hytime. Esto es lo que supone el modelo SAM, sirve ahora como punto común de referencia para las dos sintaxis, y la comparación de expresiones sintácticas se realiza por cotejo de los modelos SAM desde los que fueron creados.

Esta solución opera para nuevas sintaxis *topic map*; es decir, para cada sintaxis no normalizada es posible hallar un modo de relacionarla con aquellas que si lo están. Incluso,

⁵⁷⁰ BIEZUNSKI, M. Introduction to the *topic maps* paradigm. En Park, J. ; Hunting, S. *XML Topic maps*. Boston : Addison-Wesley, 2002. p.29.

⁵⁷¹ BIEZUNSKI, M. ; NEWCOMB, S. ; BRYAN, M. *Guide to the topic map standardization*. En: <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0323.htm>. Consultado el 15/12/2009.

proporciona un modo de mapeo desde sintaxis que no representan directamente *topic maps*, pero sí información relacionada, tal como ocurre con los lenguajes NewsML y XFML⁵⁷².

A pesar de lo que pueda pensarse, el desarrollo de la normativa *topic map* no se hizo de forma aislada de la red de redes. En el año 2000, se constituyó la organización *Topic maps.org* con el objetivo de desarrollar los *topic maps* sobre la base de las recomendaciones del consorcio W3C respecto a XML y XLink. A raíz del trabajo de la organización se publicó una versión básica de la especificación XTM ese mismo año, y la especificación oficial XTM 1.0 al año siguiente. A finales del año 2001 la organización ISO aprobó la DTD de XTM 1.0 junto con unas correcciones técnicas de la norma ISO 13250 que convertían a notación XTM la parte de la norma ISO que estaba en sintaxis XML/XLink.

Una vez editada la sintaxis XTM 1.0, *Topic maps.Org* pasó a centrarse en aplicaciones de dominio vertical, centradas en un dominio de conocimiento, y publicitar la tecnología *topic map*, ya que la labor de edición y desarrollo de normas *topic map* pasó a ser competencia exclusiva del comité SC34.

En el año 2001, se constituyeron, en el seno del consorcio OASIS, tres comités técnicos:

1. Materias publicadas para *topic maps* que desarrolla una recomendación sobre cómo definir, documentar, publicar y aplicar materias publicadas.
2. Materias publicadas *topic maps* para geografía y lenguas, para la definición de colecciones de materias públicas referidas a lenguas y objetos geográficos que aparecen en publicaciones.
3. Vocabulario para normas y tecnologías XML, donde se define un vocabulario, colección de materias publicadas, para el dominio de normas y tecnologías XML a fin de proporcionar una colección de referencia sobre *topics*, *topic types* y *association types*. Más tarde, en el año 2003 se creó un cuarto comité técnico para desarrollar un paquete de pruebas de conformidad con el propósito de evaluar los motores *topic map*.

⁵⁷² Ibid., p.2.

El subcomité SC34 comenzó posteriormente dos iniciativas de normalización:

1. *Topic map* query language (TMQL, ISO/IEC 18048). Hoy día, existen borradores de la norma y en julio de 2009 se aprobó un documento final como norma en proceso de consulta entre los desarrolladores de aplicaciones para *topic maps*⁵⁷³.
2. *Topic map* constraint language (TMCL, ISO/IEC 19756). También cuenta con numerosos borradores y se espera que se apruebe la norma para el 15 de junio de 2010⁵⁷⁴.

Como vemos en la tabla siguiente, editada por la ISO, acerca de los distintos estadios de desarrollo de las normas⁵⁷⁵, la correspondiente a TMQL se encuentra en el estadio 40.99; es decir, está en la etapa de consultas para su aprobación final. La norma correspondiente a TMCL se encuentra en el estadio 40.60 que corresponde a la próxima aprobación final de la norma.

STAGE	SUBSTAGE							
	00	20	60	90	92	93	98	99
	Registration	Start of main action	Completion of main action	Decision Substages	Repeat an earlier phase	Repeat current phase	Abandon	Proceed
00 Preliminary stage	00.00 Proposal for new project received	00.20 Proposal for new project under review	00.60 Close of review				00.98 Proposal for new project abandoned	00.99 Approval to ballot proposal for new project
10 Proposal stage	10.00 Proposal for new project registered	10.20 New project ballot initiated	10.60 Close of voting	10.92 Proposal returned to submitter for further definition			10.98 New project rejected	10.99 New project approved
20 Preparatory stage	20.00 New project registered in TC/SC work programme	20.20 Working draft (WD) study initiated	20.60 Close of comment period				20.98 Project deleted	20.99 WD approved for registration as CD
30 Committee stage	30.00 Committee draft (CD) registered	30.20 CD study/ballot initiated	30.60 Close of voting/ comment period	30.92 CD referred back to Working Group			30.98 Project deleted	30.99 CD approved for registration as DIS
40 Enquiry stage	40.00 DIS registered	40.20 DIS ballot initiated: 5 months	40.60 Close of voting	40.92 Full report circulated: DIS referred back to TC or SC	40.93 Full report circulated: decision for new DIS ballot		40.98 Project deleted	40.99 Full report circulated: DIS approved for registration as FDIS
50 Approval stage	50.00 FDIS registered for formal approval	50.20 FDIS ballot initiated: 2 months. Proof sent to secretariat	50.60 Close of voting. Proof returned by secretariat	50.92 FDIS referred back to TC or SC			50.98 Project deleted	50.99 FDIS approved for publication
60 Publication stage	60.00 International Standard under publication		60.60 International Standard published					
90 Review stage		90.20 International Standard	90.60 Close of review	90.92 International Standard to be	90.93 International Standard			90.99 Withdrawal of International

Figura 43. Cuadro resumen de los estadios de desarrollo para la aprobación de normas ISO.

⁵⁷³ ISO/IEC FCD 18048. En:

http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?ics1=35&ics2=240&ics3=30&csnumber=52119.

Consultado el 15/12/2009.

⁵⁷⁴ ISO/IEC FCD 19756. En:

http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?ics1=35&ics2=60&ics3=&csnumber=39987

Consultado el 15/12/2009.

⁵⁷⁵ International harmonized stages codes. En:

http://www.iso.org/iso/standards_development/processes_and_procedures/stages_description/stages_table.htm#s40. Consultado el

15/12/2009.

Además, el subcomité SC34 decidió definir dos modelos de datos para el paradigma *topic map*⁵⁷⁶:

1. el modelo de referencia, RM;
2. el modelo de aplicación normativo, SAM.

Entendieron que una vez publicada la sintaxis de intercambio de información *topic map* XTM 1.0, se hacía necesario el desarrollo de normas que desarrollaran distintos aspectos de la norma básica. Además, un modelo de datos preciso ayuda a los desarrolladores de software a mejorar tanto su comprensión del paradigma de los *topic maps* como a evitar malinterpretaciones del texto de la norma. En este modelo de aplicación normativo se recogen las sintaxis HyTM y XTM, así como la generación coherente de estos formatos. Esto hace que tanto TMQL como TMCL se definan de acuerdo con este modelo normativo.

Un formato canónico y un schema sobre cómo añadir direcciones también han sido añadidos en relación con el modelo normativo para homogeneizar el proceso de edición de *topic maps* entre los distintos motores *topic map* hasta ahora creados. Todas estas normas las podemos visualizar en la figura que se sigue a continuación⁵⁷⁷.

⁵⁷⁶ BIEZUNSKI, M ; NEWCOMB, S. ; BRYAN, M. *Guide to the topic maps standards*. En <http://www.y12.doe.gov/sgml/sc34/document/0323.htm>. Consultado el 15/12/2009.

⁵⁷⁷ RATH, H. *The Topic maps handbook*. Germany : Empolis, 2003. p.46.

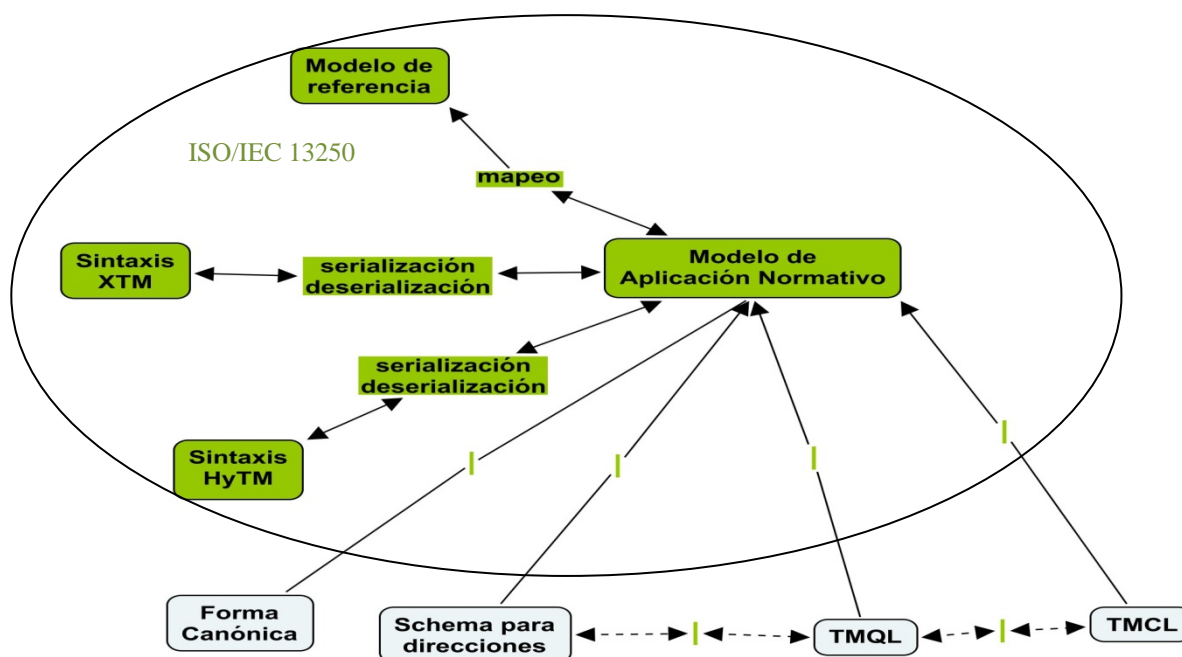


Figura 44. Esquema general de la familia de normas *Topic map*.

Tal y como vemos en el esquema, el lenguaje TMQL depende del modelo de aplicación normativo pero no está integrado en la norma ISO 13250, ya que tiene su propia norma. El objetivo de la norma ISO 18048 para los sistemas de gestión de *topic maps* es análogo al papel que el lenguaje SQL tiene para los sistemas de gestión de bases de datos relacionales; es decir, proporcionar un interfaz normalizado para la consulta, edición y actualización de *topic maps*.

TMQL⁵⁷⁸ es un lenguaje en cuya norma pueden diferenciarse dos partes: la que define cómo consultar *topic maps*, y cómo crear y actualizar *topic maps*. La norma cubre todas las aplicaciones *topic map* distribuidas en internet⁵⁷⁹, pero que alcanza también a redes de dominios particulares como la jurídica entre las bases de datos de los países miembros de la Unión Europea.

La norma ISO 19756 para el lenguaje TMCL establece un marco para la definición de schemas *topic map* para aplicaciones específicas de un dominio. Permite la validación

⁵⁷⁸ GARSHOL, L. ; BARTA, R. *Topic maps Query Language*. En: <http://xml.coverpages.org/TMQL-18048-FCD-2008-07-15.pdf>. Consultado el 15/12/2009.

⁵⁷⁹ RATH, H. *The Topic maps handbook*. Germany : Empolis, 2003. En: [http://www.sts.tu-harburg.de/~r.f.moeller/lectures/anatomie-i-und-k-system/empolistopic mapswhitepaper_eng.pdf](http://www.sts.tu-harburg.de/~r.f.moeller/lectures/anatomie-i-und-k-system/empolistopic%20mapswhitepaper_eng.pdf). Consultado el 15/12/2009.

semántica y la edición guiada de *topic maps*. Tal y como dice la propia norma, su objetivo es⁵⁸⁰:

- permitir la definición de clases de *topic maps* a fin de permitir la documentación de la estructura y semántica de una clase de *topic maps*.
- proporcionar un fundamento para la definición de aplicaciones *topic map* específicas de dominio.
- proporcionar medios de validación que aseguren la coherencia dentro de un *topic map* o entre una clase de *topic maps*.
- permitir a las aplicaciones proporcionar interfaces de usuario más intuitivos para la creación y mantenimiento de *topic maps*.
- permitir diferenciar entre tareas de modelado y poblamiento de *topic maps*.

En cuanto al modelo de datos *topic map*, debemos distinguir dos. Uno es el modelo de referencia, RM. Otro es el modelo de aplicación normativo, SAM.

El modelo de referencia, normativo desde la publicación de la norma ISO 13250-5⁵⁸¹, define los principios básicos para estructuras de declaración general de ontologías. Proporciona la base para el establecimiento del paradigma *topic map* y es independiente de cualquier sintaxis particular. Como se dice en la propia norma, proporciona una base robusta y predecible para el cotejado de conocimiento sobre materias, con independencia de la diversidad de ontologías que gobiernan la comprensión de ese conocimiento. En el borrador del modelo de referencia, un *topic map* se entiende como una colección de declaraciones en la que cada una de ellas manifiesta la existencia de tipos de relaciones entre colecciones específicas de materias de conocimiento. Como vemos en el esquema, este modelo proporciona el fundamento formal para otras normas *topic maps* como TMQL y TMCL.

⁵⁸⁰ *Topic maps Constraint Language*. En: <http://www.isotopicmaps.org/tmcl/tmcl.html>. Consultado el 15/12/2009.

⁵⁸¹ DURUSAU, P. ; NEWCOMB, S. ; BARTA, R. *Topic maps Reference Model, 13250-5*. En: <http://www.isotopicmaps.org/TMRM/TMRM-7.0/tmrm7.pdf> . Consultado el 15/12/2009.

El modelo de aplicación normativo, SAM, define el modelo de datos *topic map* en términos de una colección de elementos *topic map*⁵⁸². Este se basa en el modelo de referencia, y por tanto es más próximo a los conceptos *topic map* para su aplicación por parte de un motor *topic map*. La diferencia entre ambos modelos está en que el modelo de referencia define la estructura de un conjunto arbitrario de declaraciones y el modelo SAM define aquellos tipos de declaraciones específicas que son parte del paradigma *topic map*, define virtualmente todas las características propias de los *topic maps* como *topic names*, *topic occurrences*, etc.

5.3. Conceptos clave de los *Topic maps*.

5.3.1. Concepto de *Topic map*.

- i) Para la norma ISO/IEC 13250⁵⁸³.

Esta norma internacional proporciona una notación normalizada para la representación intercambiable de información acerca de la estructura de las fuentes de información empleadas para definir los *topics*, y las relaciones entre éstos. A una colección de uno o más documentos interrelacionados que emplean la notación definida por esta norma internacional se le denomina *Topic map*. En general, la información estructural comunicada por los *topic maps* incluye:

- grupos de objetos de información direccionables en torno a *topics* ('ocurrencias'),
- relaciones entre *topics*: *associations*.

Un *topic map* define un marco *topic* multidimensional en el sentido de espacio en el cual las localizaciones son *topics*, las distancias entre *topics* son medibles en términos del número de *topics* que deben ser visitados con el objeto de ir de un *topic* a otro, y los tipos de relaciones que definen el camino de un *topic* a otro a través de los *topics* intervinientes son muy variados. Un tutorial explicativo de lo que significan todos estos elementos para constituir un *topic map* lo vemos en la siguiente figura⁵⁸⁴.

⁵⁸² GARSHOL, L. ; MOORE, G. *The Standard Application Model for Topic maps*. En:

<http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0299.htm> . Consultado el 15/12/2009.

⁵⁸³ La norma completa se puede encontrar en: <http://www.y12.doe.gov/sgml/sc34/document/0129.pdf>. Consultado el 15/12/2009.

⁵⁸⁴ *Topic map Martial Notation*. En: <http://cd.tp/TMMN/>. Consultado el 15/12/2009.

- enlaces desde los *topics* a las fuentes de información que se consideran son ocurrencias de las materias a las que los *topic* cosifica.
- *associations* entre los *topics*.

La norma ISO/IEC 13250:2000 define y normaliza las características y propiedades de los *topic maps*. Entre estas propiedades establece el empleo de dos lenguajes de marcado, HyTime y XTM para su declaración. HyTime es un vocabulario derivado del metalenguaje SGML destinado a un uso general, mientras que XTM⁵⁸⁶ se basa en XML para permitir su empleo con otras tecnologías web. En la norma ISO se explicitan aquellos tipos de elementos que conjuntamente definen los cuatro elementos clave de los *topic maps*: *topics*, *associations*, *occurrences* y *roles*. Un *topic* no es más que un término escogido para la representación de una materia. Como nos dice Pepper, el *topic* es un símbolo empleado dentro de un *topic map* para representar una única materia para que se puedan realizar declaraciones acerca de esa misma materia⁵⁸⁷. Su declaración con la sintaxis propia permite que le sean asignadas características capaces de interrelacionar nuestro *topic* con otros y a su vez con los recursos informativos pertinentes. Por lo tanto, se hace necesario caracterizar las relaciones entre los *topics*, y es esta característica la que nos permite superar los límites de la organización de contenidos mediante la tipificación de las relaciones; es decir, la agrupación de las relaciones en clases⁵⁸⁸. Esta es una característica destacada de los *topic maps*, porque ni los métodos convencionales de modelado de datos ni los métodos de modelado semántico para la recuperación de información tales como taxonomías o tesauros han empleado la agrupación y tipificación de relaciones.

No obstante, XTM proporciona un mecanismo para la distinción de los matices individuales de cada *topic* a través de su *role*; de forma que cada uno implicado en una *association* puede decirse que juega un *role*.

En consecuencia, los *topic maps* suponen una confluencia de modelos teóricos que provienen de campos tan diversos como la inteligencia artificial, la documentación, la lingüística o la organización y representación del conocimiento y la teoría de la

⁵⁸⁶ Se trata de la contracción de XML para *topic maps*, es un schema específico del lenguaje XML.

⁵⁸⁷ PEPPER, S. Topic Maps. En BATES, M ; MACK, M. *Encyclopedia of Library and Information Sciences*. Los Angeles : CRC press, 2009. p.7. En: http://www.ontopedia.net/pepper/papers/ELIS-Topic_maps.pdf. Consultado el 15/12/2009.

⁵⁸⁸ GOLDFARB, CH. ; PRESCOD, P. *XML handbook*. Upper Sadle River : Prentice-Hall, 2002. p.642.

comunicación. Su propósito es la descripción de estructuras de conocimiento asociadas a recursos de información. Ello supone que es una tecnología para la gestión del conocimiento basado en la búsqueda por navegación sobre la base de una red de hiperenlaces entre materias. Como consecuencia, a esta tecnología se la conoce como el GPS del universo de información, siguiendo la máxima de que “un libro sin un índice es como un país sin un mapa”⁵⁸⁹.

En la navegación, como estrategia de búsqueda de información, el usuario puede realizar una búsqueda superficial, sin una idea clara de la información que desea obtener, o bien adoptar la actitud de estar ante una demanda informativa que el usuario no es capaz de satisfacer, bien por su mala formulación de la consulta o bien por el desconocimiento de la organización de las fuentes de información y de la arquitectura de los sitios web⁵⁹⁰. Para el primer tipo de usuario se le ofrece la posibilidad de emplear lenguajes documentales como clasificaciones o tesauros que les asisten en la localización de información tanto en la formulación de la pregunta como en la selección de los resultados. Para el segundo tipo de usuario se cuenta con directorios, taxonomías y mapas de sitio que permiten al usuario navegar desde un recurso de información hasta otro sin perder tiempo navegando por recursos no pertinentes a la consulta.

En esta segunda actitud del usuario, los *topic maps* juegan un papel clave al conjugar la asistencia en la formulación de las preguntas con la orientación del usuario en la navegación por una colección de recursos informativos⁵⁹¹.

Si bien es posible representar estructuras cognitivas muy complejas con los *topic maps*, su fundamento es muy sencillo. Fundamentalmente se estructura a partir de tres elementos clave, que conjuntamente son conocidos como el TAO de los *topic maps*, *topics* (T), *associations* (A) y *occurrences* (O) que como ejes de coordenadas en un mapa permiten la localización de un recurso informativo durante el proceso de navegación.

La navegación por el universo de información es demasiado compleja como para poder ser efectuada sin un mapa que nos sitúe en un contexto informativo y nos oriente hacia la

⁵⁸⁹ PEPPER, S. *The TAO of Topic Maps*. En: <http://www.gca.org/papers/xml europe2000/pdf/s11-01.pdf>, p.2. Consultado el 15/02/2010.

⁵⁹⁰ PASSIN, T. *Explorer's guide to the semantic web*. Greenwich : Manning, 2004. p.62.

⁵⁹¹ *Ibid.*, p.63.

satisfacción de nuestras demandas informativas. Esto es lo que ocurre cuando deseamos localizar una determinada materia en un libro, podemos ir ojeando todas las páginas del mismo o bien podemos acudir al índice alfabético del mismo, donde se nos dicen las páginas en las que aparecen las materias de nuestro interés. Por ello, un índice constituye una ventaja para todo texto, pero veremos que los índices de los libros son extrapolables a los recursos digitales con independencia de su formato: texto, imagen fija o vídeo. Estos índices son, en el contexto digital, un servicio de valor añadido dentro de la arquitectura de información de los sitios web. A priori, se podrían identificar estos índices con los mapas de sitio, pero el usuario necesita ser capaz de navegar por materias y entre distintos sitios con independencia de la existencia de hiperenlaces entre dos sitios que por las materias tratadas pueden estar muy relacionadas. Es aquí donde los *topic maps* recogen la idea que Bonura daba de los índices de libros como mapas para los lectores en el cual el indizador realiza una labor de cartografiado de la información⁵⁹².

Hasta ahora no ha habido una opción equivalente a los índices alfabéticos en el ámbito electrónico, aunque sí que ha existido la posibilidad de generar índices automáticamente en los procesadores de texto, estos sólo sirven para la recuperación en los documentos textuales a partir de los cuales se han generado estos índices. Sin embargo en el mundo red, sabemos que el concepto de documento es mucho más amplio que nuestra concepción de documento textual, dado que engloba documentos en formatos tan diversos como sonidos, imágenes, audiovisuales que difícilmente pueden ser indizados automáticamente con el indizador de un procesador de texto. Además, por las características intrínsecas del hipertexto, o más bien del hipermedia, la distinción entre documentos individuales desaparece. Es por ello que en el ámbito de la recuperación de información y gestión del conocimiento pasa a hacerse necesaria la capacidad de los índices para recuperar documentos interrelacionados que sean generados automáticamente a partir de corpora documentales en formatos heterogéneos. Una de las estrategias para lograrlo es mediante la fusión de los índices asociados a cada uno de los documentos que integran un corpus documental, conformando un metaíndice a partir del cual sea posible recuperar una materia simultáneamente en distintos documentos. No obstante se generan problemas importantes para el proceso de indización, especialmente para la indización automática, que debe tratar no sólo con documentos textuales sino también con documentos audiovisuales. Documentos audiovisuales que en el contexto de la web social 2.0 hacen difícil la

⁵⁹² BONURA, L. S. *The art of indexing*. New York : Wiley, 1994, p.17-18.

normalización en la asignación de descriptores incluso por parte de los autores de los documentos. Es en este contexto donde los *topic maps* cobran sentido como una nueva metodología que relaciona la tradicional indización automatizada con la organización y representación del conocimiento para generar un lenguaje controlado ubicuo capaz de situar al usuario como agente activo en la satisfacción de sus demandas informativas.

5.3.2. La norma ISO/IEC 13250.

La norma ISO/IEC 13250 se estructura en siete secciones. La primera de ellas corresponde a una visión general de los *topic maps* y sus conceptos básicos⁵⁹³. En ella se describe de manera general cada una de las partes de la norma y cómo éstas se adecúan entre sí en un todo coherente.

La segunda sección, especifica un modelo de datos para *topic maps*. Define la estructura abstracta de los *topic maps* utilizando el formalismo informativo del consorcio W3C y explica cada una de sus partes. Además, se proporcionan las reglas predeterminadas para la fusión en *topic maps*, así como algunos principios básicos sobre cuando considerar materias publicadas. Este modelo de datos, más conocido por TMDM o SAM⁵⁹⁴, es fundamental para las siguientes secciones de la norma así como para las normas de los lenguajes TMQL y TMCL.

La tercera sección se corresponde con la sintaxis XTM para *topic maps* basada en XML, XPointer e IRIs⁵⁹⁵. Las expresiones sintácticas admitidas en los documentos XTM están restringidas al empleo de un schema RELAX-NG. La semántica de esta sintaxis se define en la sección segunda de la norma.

La cuarta parte, ISO/IEC 13250-4 o más conocida por CXTM, describe cómo representar de forma canónica los datos, que no tengan más de una representación posible⁵⁹⁶. Para ello se define lo que significa transformar una instancia del modelo de datos *topic map* en una instancia del modelo de elementos info:et de XML. Posteriormente, se especifica el tipo de

⁵⁹³ NAITO, M. ; PEPPER, S. *Draft 13250-1 Topic maps—Overview and Basic Concepts*. En: www.jtc1sc34.org/repository/0877.zip. Consultado el 15/12/2009.

⁵⁹⁴ DATA MODEL. En: www.isotopic-maps.org/sam/sam-model. Consultado el 15/10/2009.

⁵⁹⁵ MASON, J. *Topic maps –XML Syntax*. En: <http://www.isotopic-maps.org/sam/sam-xtm/>. Consultado el 15/12/2009.

⁵⁹⁶ GARSHOL, L. ; NAITO, M. *ISO 13250-4: Topic maps- Canonicalization (CXTM)*. En: <http://www.isotopic-maps.org/cxtm/>. Consultado el 15/12/2009.

orden de cualquier colección de elementos informativos de ese modelo. Es muy utilizado para la evaluación de motores *topic map*.

La quinta parte define el modelo de referencia, RM, para el establecimiento de un modelo abstracto de *topic maps* que soporte la definición de lenguajes guía para modelos de más alto nivel para *topic maps*⁵⁹⁷. Para ello, se declara aquí los requisitos para estos modelos, lo cual es imprescindible para otras normas *topic map* como la correspondiente a TMQL y a TMCL.

La sexta sección es para la sintaxis compacta de *topic maps*⁵⁹⁸, CTM, como alternativa a la sintaxis compleja de XML definida en ISO/IEC 13250-3. Ambas sintaxis son idénticas en términos de significado y procesamiento de acuerdo con el modelo de datos para *topic maps* TMDM y la forma canónica CXTM. La norma recoge la experiencia de edición de lenguajes para la edición simplificada de *topic maps* respecto de XTM, tales como LTM⁵⁹⁹ o AsTMa⁶⁰⁰.

La sección séptima se refiere a cómo debe realizarse la representación gráfica de los *topic maps*, GTM⁶⁰¹. Para ello, se define una notación gráfica para la representación de *topic maps* para ser utilizada en diseño, enseñanza, o cualquier tipo de explicación que se desee emplear sin necesidad de contar con una sintaxis para *topic maps*, dado que en muchas ocasiones resulta más útil la representación no sintáctica de un *topic map*.

La relación que existe entre cada una de estas secciones para conformar la totalidad de la norma ISO/IEC 13250 se puede ver en el siguiente esquema⁶⁰². Donde la norma 13250 comprende todo aquello que está dentro de la forma trapezoidal.

Norma ISO 13250

⁵⁹⁷ DURUSAU, P. ; NEWCOMB, S. ; BARTA, R. *Topic maps Reference Model, 13250-5*. En: <http://www.isotopicmaps.org/TMRM/TMRM-7.0/tmrm7.pdf>. Consultado el 15/12/2009.

⁵⁹⁸ HEUER, L. ; HOPMANS, G. ; BOGACHEV, D. *ISO 13250-6: Topic maps- Compact Syntax (CTM)*. En: <http://www.isotopicmaps.org/ctm/>. Consultado el 15/12/2009.

⁵⁹⁹ GARSHOL, L. *The Linear Topic map Notation*. En: <http://www.ontopia.net/download/ltn.html>. Consultado el 15/12/2009.

⁶⁰⁰ BARTA, R. *AsTMa*. En: <http://astma.it.bond.edu.au/astma=-spec-xtm-1.0r1.8.dbk>. Consultado el 15/12/2009.

⁶⁰¹ GARSHOL, L. ; LEE, J. *GTM*. En: <http://www.isotopicmaps.org/gtm/>. Consultado el 15/12/2009.

⁶⁰² DURUSAU, P. ; NAITO, M. *Overview and basic concepts*. En: C:/Users/David%20Martul/Desktop/0877.htm. Consultado el 15/12/2009.

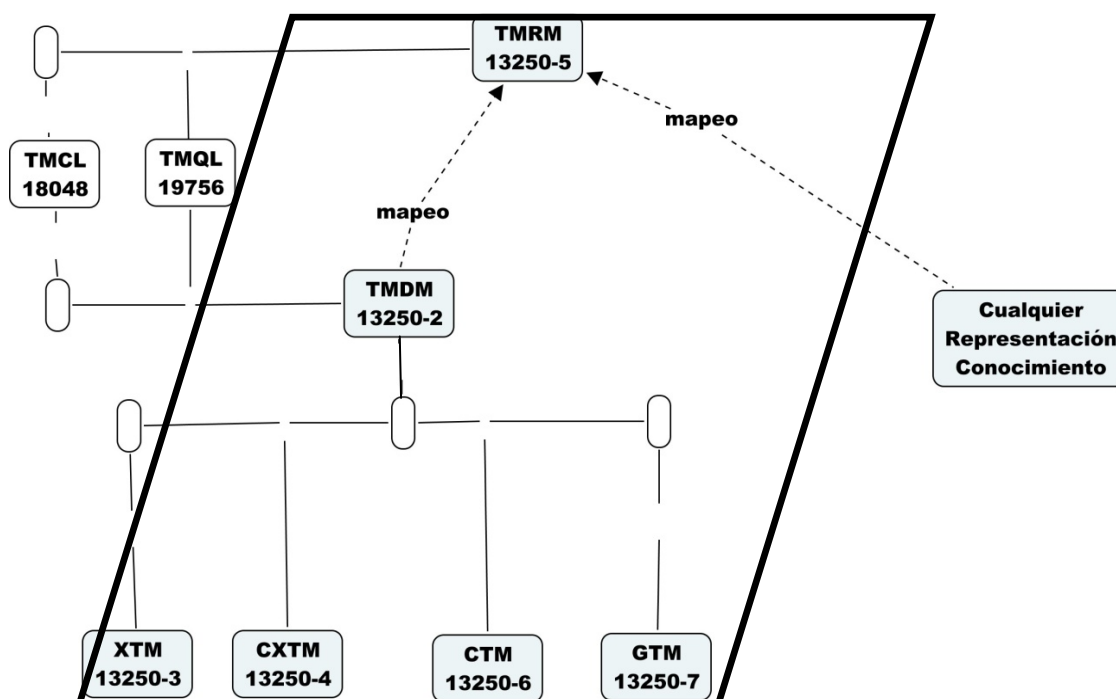


Figura 46. Esquema de las secciones constitutivas de la norma ISO/IEC 13250.

5.4. Estructura de los *Topic maps*.

Como modo de concretar el empleo de la notación *topic map* definida en esta especificación, vamos a ver el siguiente ejemplo⁶⁰³. Supongamos que deseamos recoger, a modo de aplicación, un tipo de información acerca de un documento que pueda ser incluida en el índice de materias de una enciclopedia en formato electrónico.

En este índice existen materias tales como por ejemplo John Sowa, o bien la materia correspondiente a su obra *Knowledge Representation*, o bien las ciudades de Boston y Madrid como materias. A su vez, estas localidades pueden también ser *occurrences*. Estas, pueden relacionarse con las materias de distintas maneras, según cómo se desee permitir a los usuarios localizar los documentos que satisfagan sus demandas informativas.

La enciclopedia está en formato electrónico, por lo tanto cada *occurrence* de una materia es una fuente electrónica, a la cual podemos asignar una dirección electrónica. Entendiendo

⁶⁰³ Tenemos un buen ejemplo de implementación de un *topic map* para una guía de turismo en: RATH, H. *The Topic maps Handbook*. Gütersloh : Empolis, 2003. En: http://www.empolis.com/download/docs/whitepapers/empolistopic_mapswhitepaper_eng.pdf . Consultado el 15/12/2009.

por dirección una expresión, normalmente breve, que permite a un procesador localizar una fuente. Son por lo tanto fuentes de información direccionables⁶⁰⁴.

Sowa, en cambio, es una fuente no direccionable. No es del todo un objeto digital, pero sí una entidad humana real. Con el objeto de representar el enlace entre la *occurrence* y la materia, cabría señalar a cada uno y decir “en esta localización se trata de esta materia”. Como no todas las materias son objetos digitales, no podemos proporcionar una dirección para todas las materias. En cambio, se proporciona un sustituto digital para la materia, la cual puede tener una dirección. A este sustituto se le denomina *topic* y al proceso de sustitución se le denomina cosificación. Decimos que el *topic* cosifica la materia, o que hace la materia real para el sistema cuando permite al sistema manipular, procesar y asignar características a la materia cosificada por el *topic*⁶⁰⁵. Cuando necesitamos una dirección para una materia, se proporciona la dirección de un *topic* al cual cosifica, y actúa como sustituto dentro del sistema.

Cuando una colección cuenta con un índice por materias se está proporcionando un tipo de mapa para que el usuario sea capaz de localizar los documentos a partir de las mismas. Pero si deseamos ser capaces de recuperar por las relaciones que existen entre ellas entonces estamos ante un *topic map*. Por ejemplo, los *topics* representantes de muchas de las obras de Sowa se muestran en lenguaje XTM así:

```
<topic id="knowledge representation">
  <instanceOf><topicRef xlink:href="#play"/></instanceOf>
  <baseName>
    <baseNameString>knowledge representation</baseNameString>
  </baseName>
  <occurrence>
    <instanceOf><topicRef xlink:href="#plain-text-format"/></instanceOf>
    <resourceRef xlink:href="http://users.bestweb.net/~sowa/ontology"/>
  </occurrence>
</topic>
<topic id="Conceptual structures">
  <instanceOf><topicRef xlink:href="#play"/></instanceOf>
  <baseName>
    <baseNameString>Conceptual structures</baseNameString>
  </baseName>
  <occurrence>
```

⁶⁰⁴ VATANT, B. *Topic Maps from representation to identity*. Conversation, names, and published subject indicators. En: Park, J. *XML Topic Maps. Creating and using topic maps for the web*. Boston : Addison-Wesley, 2002. pp.73-74.

⁶⁰⁵ BIEZUNSKI, M. Introduction to the Topic Maps paradigm. En: PARK, J. *XML Topic Maps. Creating and using topic maps for the web*. Boston : Addison-Wesley, 2003. p.28.


```
<instanceOf><topicRef xlink:href="#plain-text-format"/></instanceOf>
<resourceRef      xlink:href="http://users.bestweb.net/~sowa/ontology"/>
</occurrence>
</topic>
```

En tesauros e índices de materia resulta útil integrar relaciones entre las materias⁶⁰⁶; por ejemplo, si recogemos las obras *Knowledge Representation* y *Conceptual Structures* en las que, Sowa es su autor, y Ontología y Wordnet son elementos de la obra *Knowledge Representation*, puede ser útil relacionar todos estos elementos entre sí para lograr una recuperación de información más precisa. En trabajos de referencia tradicionales, este tipo de relaciones son utilizadas para guiar al compilador en la creación de referencias cruzadas. Se debe destacar que estas relaciones no aparecen entre las *occurrences* sino únicamente entre las materias en sí mismas. Una representación digital de ellas puede ser totalmente independiente de las *occurrences* y puede ser aplicada a muy diferentes colecciones de fuentes. La representación digital de relaciones entre materias toma la forma de relaciones o *associations* entre *topics* que cosifican las materias.

Una *association* que represente la relación entre Sowa y su obra *Knowledge Representation* se declara en sintaxis XTM como sigue:

```
<association>
  <instanceOf><topicRef
xlink:href="#written-by"/></instanceOf>
  <member>
    <roleSpec><topicRef xlink:href="#author"/></roleSpec>
    <topicRef xlink:href="#sowa"/>
  </member>
  <member>
    <roleSpec><topicRef xlink:href="#work"/></roleSpec>
    <topicRef xlink:href="#Knowledge Representation"/>
  </member>
</association>
```

Las *associations* expresan relaciones que son inherentemente multidireccionales⁶⁰⁷: Si “*Knowledge Representation* ha sido escrito por Sowa”, automáticamente se sigue que “Sowa escribió la obra *Knowledge Representation*”; es la misma relación expresada en distintos modos. En vez de direccionabilidad, las *associations* utilizan funciones para distinguir entre las formas en que participan sus elementos⁶⁰⁸. El ejemplo anterior puede ser serializado

⁶⁰⁶ MAZZOCCHI, F. [et al.]. Relational Semantics in Thesauri: Some remarks at theoretical and practical levels. *Knowledge Organization*, vol.34, n°4, 2007. p.199.

⁶⁰⁷ GOLDFARB, CH. ; PRESCOD, P. *XML handbook*. Upper Saddle River : Prentice Hall, 2002. p.635.

⁶⁰⁸ Id.

utilizando lenguaje natural como sigue: “Existe una relación ‘escrito por’ entre Sowa (quien ejerce el papel de autor) y *Knowledge Representation* (que ejerce el papel de ‘obra’)”.

De esta manera, no hay un límite intrínseco para los tipos de relaciones entre materias que podamos registrar. En algunas proposiciones como, “viví en” y “ejemplo de” será clara el tipo de relación establecida; para otras proposiciones serán de interés otro tipo de relaciones entre materias. Esto es así porque los *topics* y sus relaciones pueden ser descritas independientemente de las *occurrences* de cualquier colección dada de fuentes de información. Esto permite que una colección de *topics* pueda relacionarse, en distintas aplicaciones, con diferentes colecciones de fuentes de información.

Una colección de fuentes de información puede ser descrita por diferentes *topic maps*, y a su vez varios *topic maps* pueden definir distintos *topics* para la misma materia. Esto hará que en la práctica sea importante fusionar *topics* con los que declarar una misma materia. Por ejemplo, a nivel abstracto, podemos decir que nuestra enciclopedia consta de una colección de fuentes de información direccionables, cada una de las cuales puede estar localizada dentro de alguna fuente de información direccionable mayor y puede pertenecer a una o más materias. Entonces, ¿cómo repercute esto en nuestro índice de materias? Pues en él se distinguirán tres partes⁶⁰⁹:

- una colección de *topics*, cada una de las cuales sirve como sustitutivo digital para cosificar una materia que puede tener uno o más nombres.
- enlaces desde los *topics* a las fuentes de información que se consideran pueden ser *occurrences* de las materias de aquellos *topics* que cosifica (por ej: discutido en, mencionado en, descrito en).
- *associations* entre *topics* (poe ej: ejemplo de, escribió/escrito por, vivió en).

Es decir, obtenemos un *topic map* donde los *topics* se refieren a cualquier entidad sobre la que queramos pensar, o representar en formato digital. Únicamente es necesario determinar tanto si dos materias son idénticas o no, como si dos *topics* cosifican la misma materia. Sin embargo, las fuentes digitales de cualquier tipo pueden a su vez convertirse en objeto de

⁶⁰⁹ DACONTA, M. ; OBRST, L. ; SMITH, K. *The semantic web: a guide to the future of XML, Web Services and Knowledge Management*. Indianapolis : Wiley, 2003. pp.170ss.

nuestra atención y por tanto ser tratadas como *topics*. Por ejemplo, una foto descriptiva de Sowa es sólo una *occurrence* del *topic* Sowa, pero también puede ser mencionada como imagen en una historia del arte, o en una lista de formatos gráficos, o en un inventario de fuentes digitales dentro de un mismo *topic map*⁶¹⁰.

5.4.1. Topics.

Los *topics* en su sentido más genérico se definen como cualquier cosa, persona o entidad, con independencia de si existe o tenga cualquier característica específica, acerca de la cual puede decirse cualquier cosa por cualquier medio⁶¹¹. De hecho, es así como la norma 13250 define lo que es una materia, término empleado para referirse a una cosa del mundo real que el *topic* supl⁶¹². La materia vendría a ser lo que Platón definió como “idea”, de forma que un *topic* vendría a ser la sombra que la idea proyecta sobre el muro de la cueva de Platón. Por tanto el *topic* es un objeto dentro del *topic map* que representa una materia.

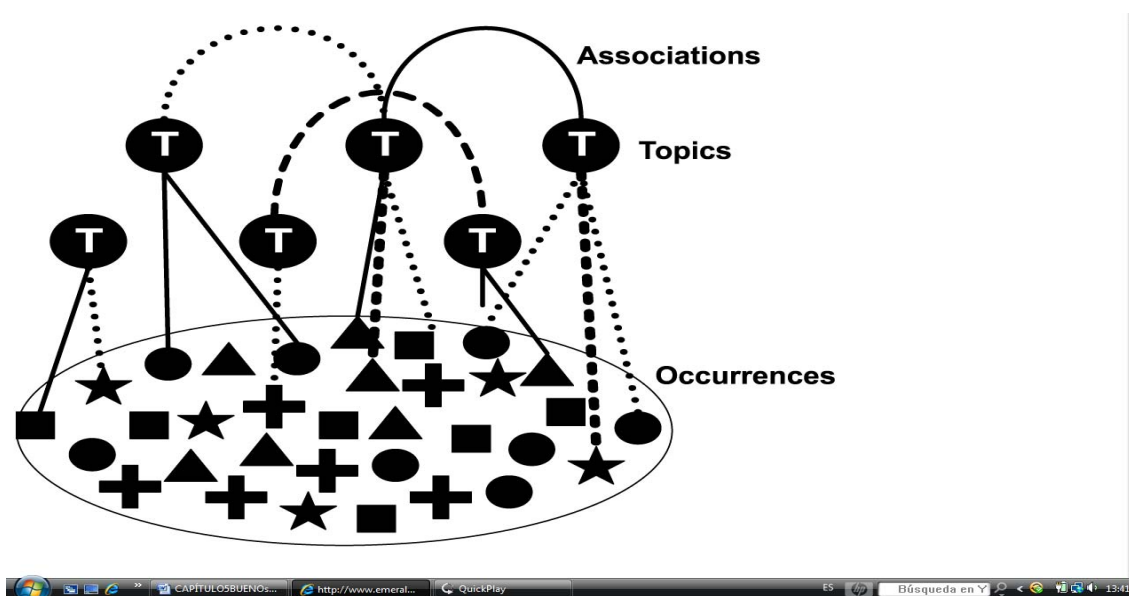


Figura 47. Representación de los *topics* en un *topic map* para un sistema gestor de conocimiento.

Según la propia norma, el corazón invisible de cada vínculo *topic* es la materia que el autor tuvo en mente cuando el *topic* fue creado⁶¹³, de forma que en cierto modo un *topic* cosifica

⁶¹⁰ GOLDFARB, CH. ; PRESCOD, P. Op. cit., p.632.

⁶¹¹ PEPPER, S. *The TAO of topic maps*. En: <http://www.gca.org/papers/xml europe2000/pdf/s11-01.pdf>, p.6. Consultado el 15/02/2010.

⁶¹² ISO/IEC 13250 *Topic Maps*. En: <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0129.pdf>, 1999, p.4. Consultado el 15/02/2010.

⁶¹³ LIU, D. ; HSU, CH. *Project-based knowledge maps: combining project mining and XML-enabled topic maps*.

una materia⁶¹⁴. Estrictamente hablando el término *topic* se refiere a aquel objeto o nodo en el *topic map* que representa la materia a la que se está haciendo referencia. Sin embargo, existe una relación individual entre un *topic* y la materia que representa, de forma que cada materia puede ser representada únicamente por un solo *topic*. En este sentido, ambos términos pueden ser empleados indistintamente. Así, en el índice alfabético de un manual de derecho los *topics* deben representar todas aquellas materias que puedan tener alguna entrada en el propio índice.

Los *topics* pueden ser categorizados de acuerdo con una tipología, de forma que cualquier *topic* es una instancia de cero o más tipos de *topic*. Esto corresponde a la categorización inherente al empleo de múltiples índices en un libro, índices de nombres, de obras, de lugares, alfabético, y al empleo de convenciones fundamentalmente tipográficas para la distinción de los distintos tipos de *topics*. Por ejemplo, “Biglino Campo” sería un *topic* del tipo “Personal” del CEPC, Madrid sería un *topic* de tipo “ciudad”. En otras palabras, la relación entre un *topic* y su tipo es una clásica relación del tipo “clase-instancia”. Exactamente lo que uno elige considerar como *topics* en una aplicación particular variará de acuerdo con las necesidades de la aplicación, la naturaleza de la información, y los usos a los cuales el *topic map* vaya destinado. Así, en un tesoro, los *topics* representarían términos, significados y dominios; en publicaciones jurídicas los *topics* representarían leyes, casos, sentencias, doctrina jurídica.

Los tipos de *topic* son declarados cada uno como un *topic* más del *topic map*, en cuyo caso las clases de *topic* “ciudad” o “personal del CEPC” son también *topics*, lo cual hace posible describir mejor las clases de *topic* en el propio *topic map*.

Los *topics* tienen tres tipos de características⁶¹⁵: *names*, *occurrences* y *roles* para las *associations*. La característica nombre de los *topics* se denomina *topic name* y constituye la designación que se otorga a un *topic*. Suele tratarse de nombres explícitos para facilitar el acceso y comprensión

En: www.emeraldinsight.com/Insight/ViewContentServlet?Filename=Published/EmeraldFullTextArticle/Articles/1720140306.html. Consultado el 15/02/2010.

⁶¹⁴ ISO/IEC 13250 *Topic Maps*. En: <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0129.pdf>, 1999, p.5. Consultado el 15/02/2010.

⁶¹⁵ BIEZUNSKI, M. Introduction to the Topic Maps paradigm. En PARK, J. *XML Topic Maps. Creating and using topic maps for the web*. Boston : Addison-Wesley, 2002. p.18.

por parte de los usuarios⁶¹⁶. No obstante, los *topics* no siempre tienen nombres. Una referencia cruzada tal como “véase la página 120” se considera como un enlace a un *topic* sin nombre explícito. Existen *names* de muchos tipos: formales, simbólicos, apodos, entradas. La norma 13250 no enumera todas las posibilidades de *names*, sino que reconoce la necesidad de que existan algunas formas que sean particularmente importantes desde el punto de vista semántico. Sin embargo, exige la necesidad de una forma normalizada de designación. Esto tiene por objeto hacer que las aplicaciones sean capaces de manejar de forma homogénea los *topics*, así como facilitar la extensión e interoperabilidad entre *topic maps*.

La norma, por tanto, facilita la asignación de múltiples nombres base a un único *topic*, y permite proporcionar variantes para cada nombre base para su empleo en contextos específicos. En la norma ISO original, las variantes fueron limitadas a dos: nombre de visualización o *display name* y nombre de clase o *sort name*. En cambio, la sintaxis XTM permite un mecanismo más general para las variantes de nombre.

La capacidad para especificar más de un nombre *topic* es empleada para indicar la aplicación de distintos nombres en distintos contextos o ámbitos tales como lengua, estilo, dominio, área geográfica o periodo histórico⁶¹⁷. Un corolario de esta característica es el esquema *topic naming constraint* para la asignación de nombres a *topics*, que establece la imposibilidad de que dos materias puedan tener exactamente el mismo nombre base en el mismo *scope*.

5.4.2. Occurrences.

Un *topic* puede ser enlazado a uno o más recursos de información que de alguna manera son considerados relevantes a un *topic* por parte del editor del *topic map*. A estos recursos se les conoce como *occurrences* del *topic*. Estos pueden ser tanto una monografía acerca de una materia concreta, o un artículo acerca de una materia en una enciclopedia, o bien una fotografía o video descrito por el *topic*, un comentario sobre el *topic* descriptivo de una

⁶¹⁶GARSHOL, L. Metadata? Thesauri? Taxonomies? *Topic Maps!* En: <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>. Consultado el 15/02/2010.

⁶¹⁷PEPPER, S. Topic Maps. En: BATES, M ; MACK, M. *Encyclopedia of Library and Information Sciences*. Los Angeles : CRC press, 2009, p.16. En: [http://www.ontopedia.net/pepper/papers/ELIS-Topic maps.pdf](http://www.ontopedia.net/pepper/papers/ELIS-Topic%20maps.pdf). Consultado el 15/02/2010.

norma jurídica en el caso de estar tratando un tema jurídico o cualquier forma en la que un *topic* pueda llegar a ser relevante para una materia concreta⁶¹⁸.

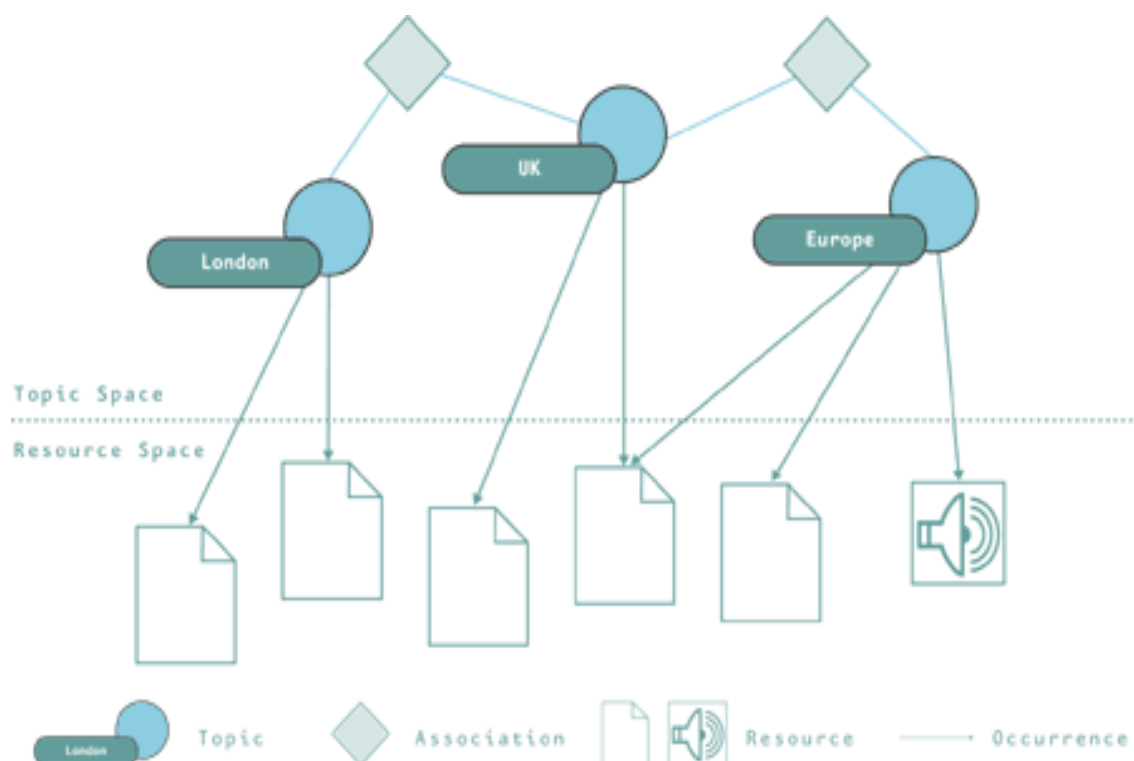


Figura 48. Representación de las *occurrences* como flechas en un *topic map*.

Las *occurrences*⁶¹⁹ son normalmente externas al *topic map*, si bien la norma no impide que estén incluidas. Se enlazan mediante el empleo de URIs en el caso de XTM o del direccionamiento con HyTime en el caso de HyTM.

Hoy día, la mayoría de sistemas para la generación manual de índices emplean alguna forma de marcado insertado en el documento a ser indizado. Una de las ventajas del empleo de *topic maps* es que ya no es necesario disponer de los documentos físicamente para que puedan ser indizados. Por otro lado, la separación en dos niveles de los *topics* y sus *occurrences* otorga una gran capacidad de organización y gestión documental a esta técnica, al diferenciar el proceso de análisis formal de los documentos del proceso de organización del conocimiento. Si en la biblioteca la organización por materias de los documentos está inextricablemente unida a su organización física, con los *topic maps* la organización de las

⁶¹⁸ PEPPER, S. *The TAO of topic maps. Finding the way in the age of infoglut*. En: <http://www.gca.org/papers/xml europe2000/pdf/s11-01.pdf>, p.8. Consultado el 15/02/2010.

⁶¹⁹ AHMED, K. ; MOORE, G. An introduction to Topic Maps. *Microsoft Architect Journal*, nº5, 2005. En: <http://msdn.microsoft.com/en-us/library/aa480048.aspx>. Consultado el 15/02/2010.

materias es independiente no sólo de la disposición física de los documentos sino incluso del espacio físico en el que se hallen.

Las *occurrences*, al igual que los *topics*, provienen de distintos tipos de fuentes: monografías, artículos, fotografías, vídeos, páginas web, etc. Esta diferenciación por clases es regulado por la norma 13250 bajo el concepto de “*occurrence role*” y de “*occurrence role type*”. La distinción entre una *occurrence role* y su clase es muy sutil, pero en general podemos decir que ambas tratan del mismo concepto sobre cómo una *occurrence* aporta información a una materia. No obstante, el *role* tiene un papel mnemotécnico en tanto que la clase, indicada sintácticamente por el atributo *type* es una referencia a un *topic* que caracteriza la naturaleza de la relevancia de la *occurrence* en relación con su materia.

En general, conviene especificar la clase del *occurrence role* ya que dota a los *topic maps* de la capacidad para ser empleados en la transmisión de más información acerca de la relevancia de una *occurrence*⁶²⁰.

5.4.3. Associations.

Hasta ahora, todos los elementos vistos han estado ligados a los *topics* como principio central de organización de la información. Los conceptos de *topic*, clase *topic*, nombre, *occurrence* y *occurrence role* permiten la organización de los recursos de información de acuerdo con un *topic* o materia, además de facilitar la elaboración de índices, pero sin llegar a explotar el potencial hipertextual de los documentos digitales. Uno de los grandes potenciales de los *topic maps* es su capacidad para establecer un amplio abanico de relaciones entre los *topics*, lo cual los convierte en un sistema de búsqueda por navegación más flexible que los tradicionales lenguajes controlados aplicados a la documentación digital⁶²¹. Para ello la norma 13250 define el elemento *topic association*⁶²².

⁶²⁰ El concepto de *occurrence role* no se encuentra definido en la sintaxis XTM dado que se le ha considerado como una estructura sintáctica característica de HyTime. El concepto de *occurrence role type* ha sido conservado en XTM, pero su denominación ha cambiado a *occurrence type* a fin de evitar su confusión con las *association roles*.

⁶²¹ Un claro ejemplo de ello es BUBL, un portal de recursos digitales que permite la recuperación a través de la navegación a partir del sistema de clasificación Dewey. En: www.bubl.ac.uk. Consultado el 01/02/2010.

⁶²² *Topic Maps*. En: http://www.wandora.org/wandora/wiki/index.php?title=Topic_Maps. Consultado el 15/02/2010.

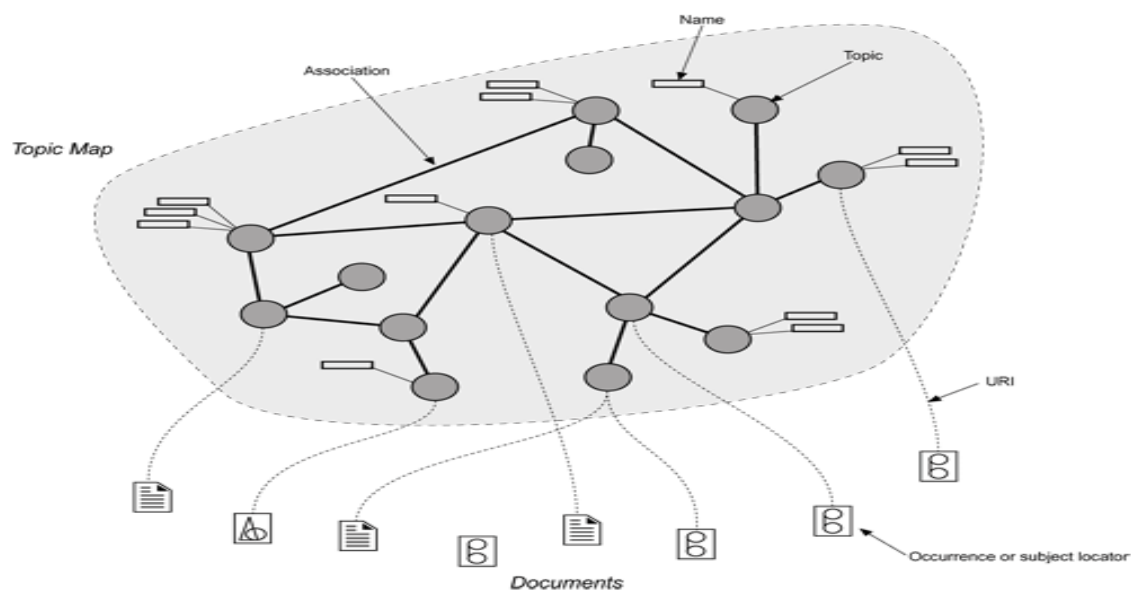


Figura 49. Las *associations* en el contexto de un *topic map*.

Un *association topic* establece una relación entre dos o más *topics*⁶²³. Por ejemplo, una sería la relación establecida entre los *topics* parlamento y ley en la sentencia: “El parlamento aprueba la ley”, donde el verbo “aprueba” es el *topic* asociación entre ambos *topics*.

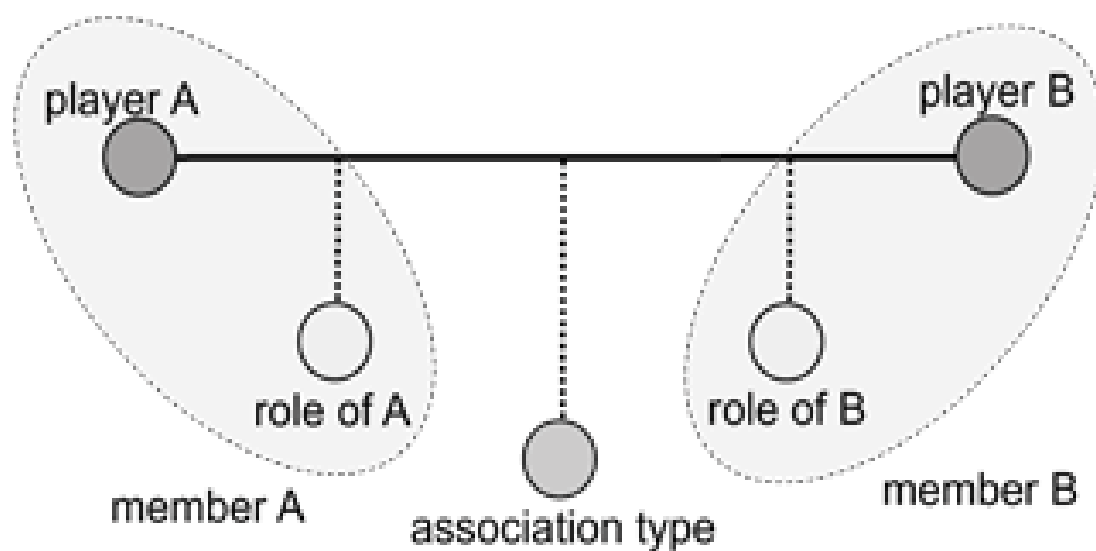


Figura 50. Representación de una *association* entre dos *topics*.

Al igual que los *topics* y las *occurrences* pueden ser agrupados en clases, también las *associations* entre *topics* pueden estar integradas en clases⁶²⁴. La clase *association*, de acuerdo con la norma

⁶²³ PEPPER, S. Topic Maps. En BATES, M ; MACK, M. *Encyclopedia of Library and Information Sciences*. Los Angeles : CRC press, 2009. p.16. En: [http://www.ontopedia.net/pepper/papers/ELIS-Topic maps.pdf](http://www.ontopedia.net/pepper/papers/ELIS-Topic%20maps.pdf). Consultado el 15/02/2010.

13250, se define como un *topic*⁶²⁵; de forma que para la relación mencionada anteriormente: “aprueba”, habría una correspondiente clase asociación que podría ser: “organismos sancionadores”.

La capacidad de establecer clases para *topic associations* incrementa la capacidad semántica de los *topic maps*, haciendo posible la agrupación de *topics* de acuerdo con un mismo tipo de relación para cualquiera de los *topics* incluidos⁶²⁶. Esto permite el diseño de interfaces muy intuitivos y usables para la navegación en grandes repositorios de información⁶²⁷.

Sin embargo, debe destacarse que las clases *topic* son consideradas como una forma especial de clase *association*, puesto que si contamos con una clase la semántica de un *topic* podría igualmente ser expresada a través de una *association*⁶²⁸, de la clase “tipo-instancia”, entre el *topic* “parlamento” y el *topic* “congreso de los diputados” en una sentencia del tipo: El congreso de los diputados es una cámara del parlamento. La razón para que haya una forma especial para este tipo de *association* obedece a las mismas razones que explican la existencia de modelos para ciertas formas de nombres, como los *surnames* o los *displaynames*. La semántica es tan general y universal que resulta muy útil normalizarla con el objeto de maximizar la interoperabilidad entre sistemas de información organizados con *topic maps*.

Asimismo, es importante destacar que a pesar de que tanto los *topic association* como las referencias cruzadas son hiperenlaces, responden a dos realidades diferentes. En una referencia cruzada, las anclas del hiperenlace se encuentran dentro de una colección de recursos de información, en tanto que con los *topic association* se está hablando de enlaces entre *topics* que son completamente independientes de cualesquier recursos de información que puedan o no existir o ser considerados como *occurrences* de estos *topics*⁶²⁹.

⁶²⁴ *Topic Maps*. En: http://www.wandora.org/wandora/wiki/index.php?title=Topic_Maps. Consultado el 15/02/2010.

⁶²⁵ ISO/IEC 13250 *Topic Maps*. En: <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0129.pdf>, p.5. Consultado el 15/02/2010.

⁶²⁶ PEPPER, S. *The TAO of topic maps. Finding the way in the age of infoglut*. En: <http://www.gca.org/papers/xmleurope2000/pdf/s11-01.pdf>, p.10. Consultado el 15/02/2010.

⁶²⁷ BATER, B. *Topic Maps. Indexing in 3-D*. En GILCHRIST, A. ; MAHON, B. *Information architecture. Designing information environments for purpose*. London : Facet, 2004. p.138.

⁶²⁸ PEPPER, S. Op. cit., p.11.

⁶²⁹ PASSIN, T. *Explorer's guide to the semantic web*. Greenwich : Manning, 2004. p.74.

Esta es una propiedad importante porque significa que los *topic maps* son activos informativos en sí mismos con independencia de si están o no conectados a documentos. Por ejemplo, si sabemos que el “congreso de los diputados” es una de las cámaras que constituyen el parlamento o que el parlamento constituye uno de los tres poderes del Estado, tanto si tenemos documentos como si no sobre estos *topics*, ya estamos proporcionando información muy útil acerca de los mismos al dejar constancia explícita de las relaciones que existen entre los mismos.

Por otro lado, debido a esta cesura entre los documentos y el *topic map*, el mismo *topic map* puede ser aplicado a varios repositorios de información como si se tratara de una red semántica portátil, al igual que distintos *topic maps* pueden ser aplicados al mismo repositorio con el fin de proporcionar distintas visiones a diferentes usuarios. Además esta separación proporciona la posibilidad de intercambiar *topic maps* entre editores, y de agregación entre ellos para constituir *topic maps* más ricos.

Cada uno de los *topics* que participa en una asociación juega un papel en la misma, a esto se le denomina *association role*. Así, en el caso de relaciones como “La cámara del Senado se encuentra en Madrid” expresada a través de una relación entre Senado y Madrid, sus *association roles* son “institución” y “lugar”. Las distintas *association roles* se agrupan en clases, y cada una de estas clases constituye un *topic*.

A diferencia de las relaciones en álgebra, las *associations* son inherentemente multidireccionales⁶³⁰. Así en *topic maps* no tiene sentido decir que A está relacionado con B pero que sin embargo B no está relacionado con A. Si A está relacionado con B, entonces B debe estar relacionado con A. Dada esta regla, la noción de *association role* adquiere mayor importancia. Un caso, puede que no sea suficiente saber que el Gobierno y la cámara del Congreso de los Diputados participan en una asociación “propuesto por”, se necesita saber cuál de los dos organismos tienen la potestad para la aprobación final de las disposiciones legislativas, y cuál tiene la potestad de iniciativa legislativa.

Esto es otro modo de aviso contra la creencia de que los nombres asignados a una clase asociación implican cualquier tipo de direccionalidad, y eso no es cierto. El Congreso de los

⁶³⁰ GOLDFARB, CH. ; PRESCOD, P. *XML handbook*. Upper Saddle River : Prentice Hall, 2002. p.636.

Diputados tiene capacidad de iniciativa legislativa pero no puede proponer una ley al gobierno para su aprobación.

5.5. *Topic maps* en la organización del conocimiento.

El objetivo de los *topic maps* es alcanzar una relación unívoca entre los *topics* y las materias que representan, con el objeto de asegurarse que todo el conocimiento acerca de una materia particular pueda ser accesible a través de un *topic*⁶³¹. Sin embargo, en ocasiones, una misma materia puede ser representada por más de un *topic*, especialmente cuando se tiene intención de fusionar dos *topic maps*. En tal caso, se hace necesario disponer de algún esquema para la definición de la identidad común de los *topics* aparentemente distintos. Así, por ejemplo, si los editores de los boletines parlamentarios de España, Francia y Reino Unido quisieran unir sus *topic maps*, entonces sería necesario poder afirmar que los *topics* Parlamento, Parlement y Parlament se refieran a la misma materia. El concepto que permite hacer esto es la identidad de materia⁶³². Cuando la materia se encuentra representada por un hiperdocumento, su identidad puede ser declarada a través de su URL. Sin embargo, la mayoría de las materias, como el concepto de poder legislativo no son directamente direccionables por medio de la URL. Este problema se resuelve por medio de indicadores de materia⁶³³. Un indicador de materia es un recurso orientado a proporcionar una indicación desambiguada de la identidad de una materia. Dado que es un documento digital, un indicador de materia tiene una dirección, normalmente una URI, que puede ser empleada como identificador de materia.

Dos *topics* que compartan uno o más indicadores de materia, o bien tienen la misma URL, en el caso de materias direccionables, o bien se les considera semánticamente equivalentes a un único *topic* que es unión de las propiedades de los dos anteriores; de forma que, en un *topic map* resultante de la unión de dos *topic maps*, un solo nodo *topic* resulta de la combinación de las características de ambos *topics*⁶³⁴.

⁶³¹ SIGEL, A. *Towards knowledge organization with Topic Maps*. En: <http://www.gca.org/papers/xml europe2000/papers/s22-02.html>. Consultado el 15/02/2010.

⁶³² PEPPER, S. *Curing the Web's Identity Crisis. Subjects indicators for RDF*. En: <http://www.ontopia.net/topicmaps/materials/identitycrisis.html>. Consultado el 15/02/2010.

⁶³³ Originalmente denominados descriptores en la primera norma ISO 13250:1999

⁶³⁴ El hecho de que dos *topics* no especifiquen la misma identidad de materia no es suficiente para concluir que los *topics* se refieran a la misma material. Lo único que prueba es que hay una identidad.

Un ejemplo de indicador de materia es un documento oficial de difusión pública. El indicador de materia publicado⁶³⁵ es un indicador de materia pensado para ser utilizado por cualquier persona y mantenido en una dirección publicitada con el propósito de facilitar el intercambio de conocimiento y su agregación, tanto a través de *topic maps* como por otros lenguajes controlados como clasificaciones o tesauros.

Las materias publicadas son un prerequisite para poder emplear la característica de portabilidad de los *topic maps*, puesto que no tiene sentido ofrecer un *topic map* a otra comunidad de usuarios si no cuenta con *occurrences* relevantes para la comunidad final de usuarios en el repositorio de documentos. Por ello, desde el consorcio OASIS se ha ido efectuando una intensa actividad de desarrollo de recomendaciones para el empleo y registro de materias publicadas⁶³⁶.

En general, es conveniente asignar metadatos a los documentos que constituyen las *occurrences* de un *topic map*⁶³⁷. Para proporcionar esta capacidad, la norma ISO 13250 incluye el concepto de faceta. Las facetas proporcionan un mecanismo para la asignación de características propiedad-valor a los documentos⁶³⁸. Una faceta, es una propiedad, y a sus valores se les denomina “valores faceta”. Las facetas son normalmente empleadas para proporcionar aquellos metadatos que no se proporcionan en los atributos de los lenguajes de marcado SGML o XML, o en un sistema de gestión de contenidos. Estos metadatos declaran propiedades tales como: lengua, seguridad, nivel de usuario, aplicabilidad, etc.

Una vez estas propiedades han sido asignadas, pueden ser empleadas para crear filtros para las consultas que produzcan subcolecciones restringidas de documentos, tales como documentos escritos en lengua francesa, o documentos donde el nivel de usuario es alumno del Diploma de Estudios Políticos y Constitucionales.

Es importante no confundir las facetas con el contexto. Las facetas no son empleadas para calificar los objetos a nivel de *topics*, sino que su propósito es añadir atributos a los

⁶³⁵ PSI, originalmente denominado descriptor público.

⁶³⁶ OASIS. En: <http://www.oasis-open.org/committees/tm-pubsubj/>. Consultado el 02/02/2010

⁶³⁷ MAICHER, L. *Autonomous topic maps: Overview*. En: <http://www.informatik.uni-leipzig.de/~maicher/MWP/ATMIndex.htm>. Consultado el 10/02/2010.

⁶³⁸ ISO/IEC 13250 *Topic Maps*, 2002. En: <http://www.informatik.uni-leipzig.de/~maicher/MWP/ATMIndex.htm>. Consultado el 02/02/2010.

documentos. En cierto modo, las facetas son una proyección desde el nivel del repositorio al nivel de los *topics*, de forma que proporciona un mecanismo útil que complementa y extiende de forma significativa la capacidad de los *topic maps*, pero en ningún caso son *topics*⁶³⁹.

El modelo *topic map* permite que se pueda decir tres cosas esenciales acerca de cualquier *topic* particular: Qué nombres tiene, qué *associations* puede tomar, y cuáles son sus *occurrences* asociadas. Estos tres tipos de afirmaciones son conocidas de manera colectiva como características *topic*⁶⁴⁰.

La asignación de características *topic* se hace siempre dentro de un contexto específico, explícito o no; de forma que no es lo mismo hablar de “poder” en Derecho que en ciencias políticas. Por tanto, el contexto es una realidad muy presente en todo momento en la recuperación de información. Según Sowa, una oración se compone de seis tipos diferentes de información; cuatro de los cuales, tiempo y modo, presuposición, enfoque y connotaciones emocionales, están relacionados de un modo u otro con el contexto⁶⁴¹. El usuario tiene una capacidad especial para tratar con el contexto, para distinguir el sentido entre discursos sintácticamente similares. Según Moreiro⁶⁴², “el lector interpreta la situación que el texto le ofrece desde su experiencia sintáctica, semántica y pragmática”. Sin embargo, los ordenadores no son capaces de distinguir los matices semánticos, hecho por el cual los *topic maps* tienen en cuenta la asignación de una característica a cada *topic*, sea un nombre, un suceso o una función adecuada a los límites del dominio, explícitos o no. A esta asignación se la conoce por contexto de la aplicación. Este contexto se define en términos de temas, y un tema se define como un elemento de una colección de *topics* empleado para especificar el campo de aplicación. En otras palabras, un tema es un *topic* que se emplea para limitar la validez de una colección de asignaciones⁶⁴³. Así por ejemplo, el nombre de la persona que ocupe el cargo de presidente del gobierno debe ser asignado a dos *topics* diferentes en ámbitos definidos por los temas “ejecutivo” y “legislativo” dado que por un lado es miembro del poder ejecutivo y por otro es miembro del poder legislativo. De este modo, se

⁶³⁹ GARSHOL, L. *Metadata? Thesauri? Taxonomies? Topic Maps. Making sense of it all*. En: <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>. Consultado el 10/02/2010.

⁶⁴⁰ Id.

⁶⁴¹ SOWA, J. *Conceptual Structures*. Reading : Addison-Wesley, 1984, p.147ss.

⁶⁴² MOREIRO GONZÁLEZ, J.A. *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural*. Gijón : Trea, 2004. p.38.

⁶⁴³ GOLDFARB, CH. ; PRESCOD, P. Op. cit., p.637.

reduce cualquier tipo de ambigüedad y se minimizan las posibilidades de error cuando se fusionan *topic maps*.

De hecho, un uso coherente del dominio de aplicación en *topic maps* no sólo es eficaz en la reducción de la ambigüedad sino como ayuda para la navegación. Por ejemplo, alterando de forma dinámica la visión del *topic map* en función del perfil de cada usuario y el modo en que se visualiza. Así pues, cualquier usuario que declare un interés concreto acerca de un cargo político en el ejercicio de su puesto en una cámara legislativa frente a su posición como miembro del gobierno puede obtener consiguientemente el nombre de esa persona ordenada por sus actuaciones en la cámara legislativa.

Análogamente, si nuestro sistema editor del *topic map* conoce cualquier característica acerca del perfil de un usuario, debe tenerlo en cuenta como presupuestos que pueden afectar al comportamiento del mapa. Si tenemos un *topic map* dedicado a la presentación de las actividades en una cámara del parlamento, el dominio de aplicación debe ser empleado para la calificación de *topics*, de forma que los distintos puntos de vista sobre la información puedan ser presentados a los distintos perfiles de usuarios con interés en la actividad parlamentaria de una de sus cámaras: ciudadanos, prensa, diputados, jueces, etc.

Asimismo, el dominio de aplicación puede ser empleado para la determinación de la denominación a asignar a un *topic* recién introducido. Así, una relación entre Negrín (presidente del gobierno) y Valencia debe ser marcada como “ejerció como presidente de gobierno desde” en el ámbito de la clase función asociación “gobierno” (Persona: Negrín) y “actuó desde” en el ámbito de la clase función asociación “localización” (Ciudad: Valencia).

Por tanto, el dominio de aplicación no debe ser confundido con las facetas. Ambos mecanismos deben ser diferentes y complementarios. Mientras que el dominio de aplicación puede ser visto como un mecanismo de filtrado basado en las propiedades de los *topics*, las facetas proporcionan filtrado sobre la base de la descripción documental de los recursos de información⁶⁴⁴.

⁶⁴⁴ Para una descripción más profunda del significado del ámbito de aplicación (scope) véase la obra de PEPPER, S. ; GRØNMO, G. *Towards a General Theory of Scope*, 2001. En: http://www.ontopia.net/topic_maps/materials/scope.htm. Consultado el 20/02/2010.

Existe una estrecha vinculación entre la organización conocimiento y los *topic maps*⁶⁴⁵. El mecanismo por el cual los *topic maps* constituyen una herramienta para la organización del conocimiento parte de las siguientes premisas:

- Los *topic maps* son un metaíndice, y por tanto en sí mismos ya tienen una capacidad para la organización del conocimiento.
- A partir de un *topic map* se puede extraer distintos tipos de lenguajes documentales, siendo uno de ellos las taxonomías empleadas para la organización del conocimiento.
- Al tratarse de una herramienta estrechamente ligada a los metadatos, posibilitan la elaboración de herramientas para la gestión del conocimiento, especialmente en Internet.
- Los *topic maps* mejoran considerablemente la recuperación por materias en documentos con formato electrónico ya que permite al usuario navegar entre las materias por las que desea recuperar y no tener que ir navegando por las sucesivas recuperaciones que vaya realizando hasta alcanzar la búsqueda deseada.

Para demostrarlas, es necesario explicar cómo los *topic maps* son una herramienta para la organización, ya no de los documentos de una organización sino del conocimiento de la misma, de forma que el documentalista no tiene porqué trabajar en la organización de la documentación de su centro sino en la organización del conocimiento. Por esta razón, es necesario elaborar unos identificadores de materia estables y coherentes con los que se pueda hacer un uso fiable y lo más amplio posible en la designación de las materias para su identificación por parte de los sistemas gestores⁶⁴⁶.

La aplicación de los *topic maps* surge en el contexto de la necesidad que las grandes corporaciones tienen de organización del conocimiento. Así, se entiende ésta como un modo de generar, codificar, transferir y preservar el conocimiento de una organización con el objeto de poder emplear de forma exhaustiva y precisa tanto la información de la

⁶⁴⁵ SIGEL, A. Topic Maps in Knowledge Organization. En: PARK, J. *XML Topic Maps. Creating and using topic maps for the web*. Boston : Addison-Wesley, 2002. pp.384-385.

⁶⁴⁶ PEPPER, S. The case for published subjects. En:

http://www.ontopia.net/topicmaps/materials/The_Case_for_Published_Subjects.pdf. Consultado el 10/02/2010.

empresa como el conocimiento de todos y cada uno de sus empleados⁶⁴⁷. Según Stewart, “el conocimiento crea riqueza⁶⁴⁸”, “el éxito llega a las empresas que tienen la mejor información o la utilizan mejor⁶⁴⁹”. Sin embargo, lo más interesante es cuando nos define el concepto de capital intelectual: “el capital intelectual es conocimiento, información y experiencia, que pueden utilizarse para crear riqueza⁶⁵⁰”. Y es en este contexto en el que los *topic maps* surgen como instrumento para codificar el conocimiento. El creciente flujo de información hace fundamental para las instituciones y organizaciones jurídicas encontrar un modo efectivo para el tratamiento, no sólo con el exceso de información sino también con el conocimiento inherente a ella. Una parte importante de este conocimiento surge del trabajo intelectual de análisis y resolución de problemas cotidianos en ellas por parte de sus miembros. De hecho, cuentan con documentalistas que acumulan y crean conocimiento compartiéndolo con otros miembros de su organización o con comunidades de interés.

El objetivo de la organización del conocimiento es conseguir que las instituciones actúen de forma tan inteligente como sea posible para asegurar su viabilidad y éxito global⁶⁵¹. Por tanto, se refiere a los métodos y herramientas para la captura, almacenamiento, organización y difusión del conocimiento dentro y entre comunidades. Según Sigel⁶⁵², la organización del conocimiento tiene que asegurarse de que todos los activos y flujos de conocimiento sean conocidos, utilizados e incrementados con el objeto de contribuir a largo plazo a incrementar el valor de los servicios ofrecidos por la institución.

El problema que se plantea para cualquier centro de documentación, especialmente del ámbito jurídico por la especialización de sus fuentes, es que resulta fácil proporcionar acceso a los datos pero es mucho más difícil a la información y al conocimiento. La cuestión está en cómo proporcionar acceso a los usuarios de una institución acceso al conocimiento de su colección especializada. La organización del conocimiento implica conectar a unas personas con otras y a estas con el conocimiento. El objetivo que se busca es que la tecnología de la información permita que el conocimiento de un individuo o

⁶⁴⁷ PEPPER, S. Towards seamless knowledge . Integrating public sector portals in Norway. En: www.ontopia.net/topicmaps/materials/Towards%20Seamless%20Knowledge.ppt. Consultado el 15/02/2010.

⁶⁴⁸ STEWART, T. *El conocimiento crea riqueza*. 2003. Consultado en El País el 15/12/2009. p. 22.

⁶⁴⁹ Id.

⁶⁵⁰ Id.

⁶⁵¹ GARCÍA JIMÉNEZ, A. *Organización y gestión del conocimiento en la comunicación*. Gijón: Trea, 2002.

⁶⁵² SIGEL, A. Towards Knowledge Organization with Topic maps. En *XML Europe 2000*, 2001. En <http://www.gca.org/papers/xml europe2000/papers/s22-02.html>. Consultado el 09/02/2010.

grupo de individuos pueda ser extraído y estructurado para ser usado por otros miembros de la institución⁶⁵³.

Asimismo, debemos ser conscientes que no basta con proporcionar una tecnología capaz de facilitar un acceso exhaustivo y preciso a la información demandada por la comunidad de usuarios. Los miembros de un organismo jurídico suelen preferir el acceso a información local, si ésta es fácilmente localizable, que no acceder a información más distante aunque sea más relevante. Esto nos lleva a plantear la cuestión de cómo mejorar la organización de la información en los sitios web de los organismos, dado que muchas veces en las grandes organizaciones se llevan procesos repetitivos sólo porque la información no es fácilmente localizable. Incluso, si la información es externa al centro, puede resultar fácil su localización para un experto en la materia pero no para el común de los usuarios de la misma por razones no sólo técnicas sino sociales y culturales que están detrás de la difusión del conocimiento. Que un autor quiera compartir su conocimiento jurídico con otras personas no sólo supone contar con unos usuarios, con su espacio y su tiempo, sino que se necesita un marco organizativo integrado en una cultura y un estilo propio promotor de la comunicación y difusión de la información jurídica⁶⁵⁴.

Muchas organizaciones cuentan con comunidades de interés compuestas por colegas en las cuales el conocimiento es compartido. Estas redes, a menudo proporcionan información de calidad sobre un área común de interés. Las comunidades de interés tienen la ventaja sobre la información digital en web de contar con información con una baja tasa de obsolescencia⁶⁵⁵. Que la información digital cuente con un elevado grado de obsolescencia supone una dificultad añadida para conseguir un apoyo técnico apropiado para las redes de información. Cuando estas redes acumulan suficientes conocimientos comunes entre ellas, es posible que sean capaces de compartir índices cuya convergencia genere nuevo conocimiento. Su viabilidad lo permite la tecnología *topic map*, por su carácter de metaíndice sobre el que se posibilita la navegación.

⁶⁵³ OBRST, L. ; LIU, H. Knowledge Representation, Ontological Engineering, and *Topic maps*. En PARK, J. *XML Topic maps. Creating and Using Topic maps for the Web*. Boston: Addison-Wesley, 2002. pp.104-105.

⁶⁵⁴ MACIÁ, M. Manual de documentación jurídica. Madrid : Síntesis, 1998. pp.21ss.

⁶⁵⁵ PLAZ LANDAETA, R. ; GONZÁLEZ AURE, N. La gestión del conocimiento organizativo. Dinámicas de agregación de valor en la organización.

En:www.mityc.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/357/05_ReinaldoPlaz_357.pdf. Consultado el 10/02/2010.

Una vez que las estructuras de conocimiento y las redes de especialistas con ese conocimiento han sido localizadas en una organización, para que la solución técnica al almacenamiento y acceso a la información sean posibles, y se cuente con una cultura de difusión del conocimiento, entonces podrá comenzarse una codificación para que los contenidos con los que trabajan puedan ser procesables por los ordenadores; es decir, el propósito de la codificación es hacer accesible el conocimiento jurídico de una institución a todos aquellos miembros que lo demanden⁶⁵⁶. El conocimiento codificado tiene muchas formas, formatos y distintas escalas en su estructuración. Puede haber desde conocimiento muy estructurado, como una taxonomía o un tesoro, hasta conocimiento menos estructurado, como una simple relación de nombres. Sin embargo, es imposible codificar todo el conocimiento de una institución⁶⁵⁷. Por ello, es aconsejable una selección de la información que pueda ser relevante para las demandas informativas de su comunidad, de forma que no se realice una selección del conocimiento útil a priori, sino una cuidada consideración de todas las materias.

El conocimiento, una vez seleccionado y codificado, debe ser estructurado. La cuestión clave en organización del conocimiento es saber cómo se estructura. La estructuración condiciona en buena parte si el conocimiento es accesible o no. La estructura facilita al usuario las tareas de recuperación respecto de su área de interés y aquellas relacionadas. Un ejemplo de esto lo representan los mapas de sitio. Estos nos permiten indicar dónde están los documentos de una materia sin necesidad de acceder a ellos, indicando a los usuarios dónde pueden localizar el conocimiento y no sólo los datos. El objetivo no es tanto proporcionar a los usuarios una información capaz de resolver sus necesidades cognitivas como de orientarlo hacia el conocimiento que está buscando, llevándole tanto a una red de documentos que lo contextualizan como a sistemas expertos capaces de resolver problemas sobre una materia⁶⁵⁸.

⁶⁵⁶ THOMAS, PH. ; KNOWLES, J. *How to use a law library. An introduction to legal skills*. London : Sweet & Maxwell, 2001. p.13.

⁶⁵⁷ Ibid., p.14.

⁶⁵⁸ CODINA, LL. ; PEDRAZA-JIMÉNEZ, R. ; ROVIRA, C. Ontologías y sistemas de información documental. En CODINA, LL. ; MARCOS, M.C. ; PEDRAZA, R. *Web semántica y sistemas de información documental*. Gijón : Trea, 2009. pp.56-57.

5.5.1. Los *Topic maps* en la organización conocimiento.

Actualmente, se está tendiendo al empleo de ontologías como herramientas para la organización del conocimiento en dominios específicos. De hecho, el objeto de la ontología⁶⁵⁹ es el estudio de las categorías de cosas que existen o pueden existir en un dominio y que son procesables por un ordenador de forma que el resultado es un catálogo de los tipos de cosas que existen en un dominio de interés D desde la perspectiva de una persona que emplea un lenguaje L con el objeto de poder hablar acerca de D.

Como podemos ver, una ontología es la definición de un dominio de conocimiento al que se puede implementar un lenguaje concreto para la organización de los elementos del dominio. Por otro lado, sabiendo que un principio básico de la organización del conocimiento es que existen dominios en los cuales el empleo de vocabularios controlados es esencial para la recuperación y tratamiento de la información. Esto explica la extremada utilidad de las ontologías como herramienta en la gestión del conocimiento en organizaciones. Sin embargo, una ontología puede ser, de acuerdo con la definición ofrecida cualquiera de los lenguajes documentales hoy en día empleados. De ahí el renovado interés que en organización del conocimiento se tiene por los tesauros, las taxonomías y sobre todo por los índices, ya que la automatización de los lenguajes documentales y su aplicación a determinados dominios de conocimiento están haciendo posible una recuperación precisa y exhaustiva por materias. No se tiene más que pensar en el uso que en dominios como el Derecho se hace de los tesauros especializados o las clasificaciones facetadas, especialmente útiles para la recuperación en sus bases de datos.

Una buena estructura de acceso por materias exige la elaboración de metaestructuras de conocimiento o metaíndices con el objeto de proporcionar los puntos de acceso por materias más pertinentes para distintas comunidades de usuarios, para ello los indizadores deben ser capaces de juzgar las potenciales materias a asignar a un registro para poder recuperarlo desde distintos puntos de vista. De ahí la estrecha relación que existe entre los metadaíndices y las herramientas para la organización del conocimiento.

⁶⁵⁹ SOWA, J. *Building, Sharing, and Merging Ontologies*. En: <http://www.jfsowa.com/ontology/ontoshar.htm>. Consultado el 19/02/2010.

Y, ¿cuál es la relación que se puede establecer entre la aplicación de las ontologías para la organización del conocimiento y la aplicación de los *topic maps* con el mismo fin?, o de otra manera, ¿cuál es la relación entre las ontologías y los *topic maps* desde el punto de vista de la gestión del conocimiento? Los *topic maps*⁶⁶⁰ son muy apropiados para la representación de ontologías ya que su propiedad clave en muchas aplicaciones es la representación del conocimiento del mundo real. El potencial de los *topic maps* reside en enlazar fuentes de cualquier procedencia en el marco de la Web Semántica, y luego organizarlas en torno a una única ontología. Asimismo, el creciente repertorio de técnicas para simplificación, fusión e interrelación de ontologías pueden ser empleadas para combinar o articular *topic maps* representando diferentes ontologías. De este modo, es posible que colecciones dispares de fuentes de información puedan ser empleadas de un modo controlable y a distintas escalas.

En líneas generales, se pueden resumir las ventajas que la organización del conocimiento obtiene de los *topic maps* en los siguientes puntos⁶⁶¹:

1. Mejora de las tasas de calidad de contenidos en procesos de indización y resumen ya que las metaestructuras de conocimiento pueden ser expresadas de forma independiente de las fuentes originales.
2. En lenguajes documentales que deban ser actualizados periódicamente, como los tesauros, los *topic maps* permiten descentralizar las estructuras de conocimiento y por tanto permite su creación y mantenimiento colaborativo, ya que deja de centrarse en un producto jerárquico y monolítico para pasar a construir una estructura en que las relaciones asociativas son las primordiales y por tanto la estructura del lenguaje documental pasa a ser mucho más flexible tanto para la interrelación de materias como para su recuperación.
3. Debido a esta flexibilidad, se permite la elaboración de metaestructuras de conocimiento mucho más complejas que permiten la recuperación de información dentro de una más sutil variedad de contextos.
4. Con el creciente número y tamaño de fuentes de información interrelacionadas, la escalabilidad se hace vital, y los *topic maps* tienen la capacidad de manejar índices de distinto escala.

⁶⁶⁰ WRIGHTSON, A. *Topic maps and Knowledge Representation*. En: http://www.ontopia.net/topic_maps/materials/kr-tm.html. Consultado el 15/02/2010.

⁶⁶¹ SIGEL, A. *Topic Maps in knowledge organization*. PARK, J. *XML Topic Maps. Creating and using topic maps for the web*. p.427.

5. Además los *topic maps* permiten la construcción de índices orientados a distintas comunidades de usuarios con los mismos elementos que componen el índice. Por tanto, se favorece una indización y una recuperación (a través del facetado y los ámbitos) más adaptable a los diferentes grupos de usuarios.
6. Estrechamente ligado con la generación, difusión y mantenimiento colaborativo de estructuras conceptuales está la necesidad de buscar y navegar dentro de estructuras semánticas que hagan el acceso a la información mucho más intuitivo.
7. Con los *topic maps* se hace fundamental dar prioridad a las relaciones asociativas frente a la tradicional hegemonía de las relaciones jerárquicas en todos los lenguajes documentales, ya que en *topic maps* no son tanto los *topics* o conceptos los que producen conocimiento como las relaciones y sus tipos que entre éstos se establecen. Es más los tipos de relación también son *topics*, de forma que, por ejemplo para un libro, se puede sacar un índice no sólo onomástico, otro alfabético, otro de lugares, otro toponímico sino que también podremos sacar un índice de inferencias o relaciones que haya entre el índice onomástico y el alfabético. Ésto permitirá la combinación simultánea mediante reglas de inferencia de varios índices, y por extensión de distintos lenguajes documentales, en la organización y acceso al conocimiento.

5.5.2. Elementos constitutivos de la organización del conocimiento con topic maps.

5.5.2.1. La relación entre materias y topics.

Una materia puede ser una persona, un objeto o un documento, pero también puede tratarse de una clase de individuos, de objetos, de documentos o un conjunto de documentos abstractos. Por tanto una materia puede ser cualquier parte del discurso que un autor desee identificar, denominar, representar o sobre la que desee realizar cualquier tipo de afirmación. De hecho en la norma ISO13250 se define materia como cualquier cosa, independientemente de si existe o tenga cualquier característica específica, acerca de la cual podamos afirmar cualquier cosa con cualquier significado⁶⁶².

⁶⁶² ISO/IEC 13250:2000 *Topic Maps. Information Technology, Document Description and Processing Languages*. Disponibles las dos ediciones en: <http://www.y12.doe.gov/sgml/sc34/document/0322.htm>. Consultado el 16/02/2010.

Las aplicaciones tratan con materias a través de representaciones formales empleando símbolos como sustitutos. De acuerdo con la terminología de los *topic maps* el término *topic* se emplea para representar estos símbolos. Un *topic* es por tanto la representación no ambigua de una única materia.

En *topic maps*, el término “fuente” se emplea en un sentido restringido como fuente de información recuperable en red, lo cual deja claro el empleo de los *topic maps* para la recuperación de documentos electrónicos.

5.5.2.2. La identificación de materias.

El objetivo de cualquier aplicación que agregue información, sea un simple índice, un sistema de clasificación bibliotecaria, un *topic map* o cualquier otro tipo de aplicación, es el de alcanzar la colocación objetiva; esto es, proporcionar puntos de unión desde los cuales todo lo que es conocido acerca de una materia dada puede llegar a ser recuperable. En *topic maps*, los puntos de unión toman la forma de *topics*. Cada *topic* debe representar exactamente una materia y cada materia debe ser representada exactamente por un sólo *topic*.

Cuando agregamos información, comparamos ontologías, o superponemos vocabularios, es crucial conocer cuando dos *topics* representan la misma materia, con el fin de ser capaces de combinarlos en un único *topic*. Alcanzar esto, suponer dejar clara la correspondencia entre un *topic* y la materia que representa. Sin embargo, esto supone que las materias sean identificadas de una manera nada ambigua.

La identificación de materias no es sólo fundamental para aplicaciones *topic map* individuales y para la interoperabilidad entre aplicaciones *topic map*; sino que es también fundamental la interoperabilidad entre aplicaciones *topic map* y otras aplicaciones que hagan un uso explícito de las representaciones abstractas de materias.

5.5.2.3. La direccionabilidad de las materias.

El método más extendido hoy en día en identificación de materias para entornos en red es a través del empleo de las direcciones, y más particularmente a través del empleo de las URI's. Este es un método adecuado cuando una materia tiene asignada una dirección bien

definida como puede ser el caso de la norma ISO/IEC 13250:2000 que tiene asignada esta dirección: <http://www.y12.doe.gov/sgml/sc34/document/0323.htm>.

En teoría, cualquier fuente de información puede ser recuperada y mostrada para su interpretación por parte del usuario. Además, su dirección puede ser empleada como un único identificador por parte de los ordenadores con el propósito de lograr la unión automatizada de materias, ya que cuando dos fuentes de información tienen la misma dirección, entonces son la misma cosa. La dirección de una materia que pase a ser una

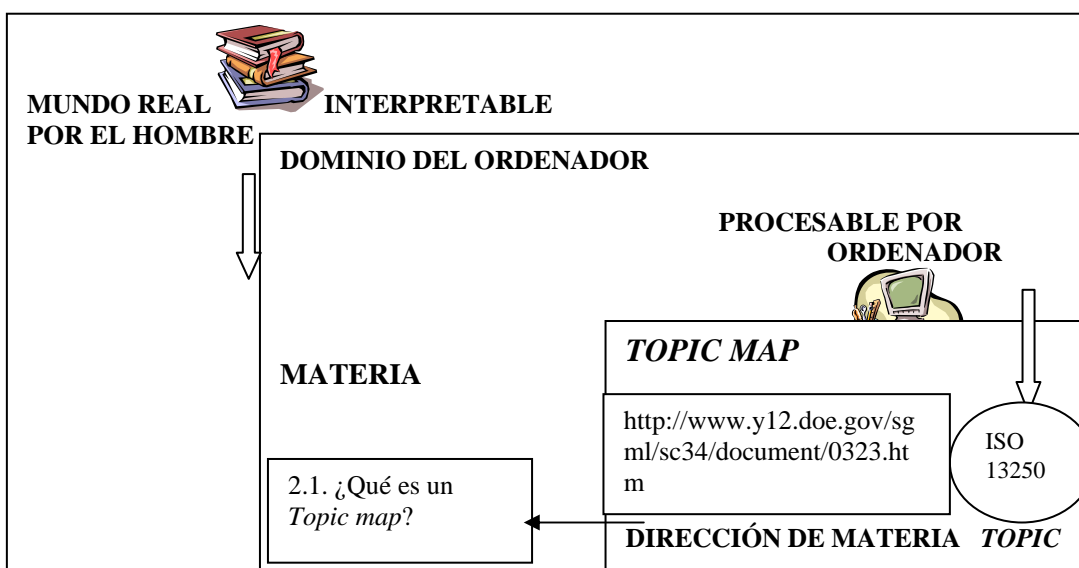


Figura 51. Esquema del *topic map* como organizador del conocimiento.

En el esquema que se ha representado en la figura, la materia sobre la cual estamos realizando una afirmación es una fuente de información, el documento que pregunta sobre lo que es un *topic map*. Por lo tanto, se tiene una red de direcciones que pueden ser empleadas como identificadores únicos por los *topics* que representan sus respectivas materias. Cuando se trata de fusionar *topics*, una aplicación es la que compara las direcciones de las materias a las que representan estos *topics*: Si dos *topics* tienen la misma dirección de materia entonces representarán la misma materia y serán fusionados.

El problema de la fusión de materias publicadas se debe a que éstas están diseñadas para poder ser direccionadas. En el paradigma *topic map*, se distingue entre materias direccionadas y materias no direccionadas. La identidad de una materia direccionada se establece desambiguadamente empleando su dirección. La identidad de una materia no direccionada se establece empleando indicadores de materia. Un indicador de materia es

simplemente una fuente que de algún modo supone una indicación de la identidad de una materia a un usuario. Por ejemplo⁶⁶³, en el caso de “Ontopia” podría ser una página HTML explicando que Ontopia es una compañía localizada en Noruega que proporciona tecnología y servicios basada en *topic map*. Esa página HTML es sólo de uso por el hombre y no por un ordenador. Es más, dado que es una fuente, la página HTML tiene una dirección, una URI, y esa URI puede ser usada por un ordenador porque es una cadena que puede ser comparada con otras cadenas. La URI de un indicador de materia es conocido como identificador de materia y es usado por los ordenadores para establecer cualquier materia.

Destacar que la misma URI puede ser empleada tanto como una dirección de materia como un indicador para diferentes materias. Por ejemplo, para la página HTML mencionada antes, la URI www.ontopia.net/about sería ambos la dirección de materia de la página de Ontopia “About” y el identificador de materia de Ontopia por sí mismo. Cualquiera puede declarar un indicador de materia para cualquier materia que queramos. Además, una única materia puede también tener múltiples indicadores de materia.

La distinción entre materias direccionables y no direccionables es uno de los muchos signos que el paradigma *topic map* lleva a la organización del conocimiento y de la información, y forma la base de la noción de materias publicadas, que se cree puede resolver el problema de la identidad como existe hoy, en particular de la World Wide Web.

5.5.2.4. Indicadores e Identificadores de materia.

Muchas veces existen materias que no son fuentes de información y por lo tanto carecen de direcciones claras y precisas. Por tanto, la identidad de tales materias no direccionables puede ser establecida sólo indirectamente a través de los indicadores de materia.

Un indicador de materia es una fuente de información que proporciona algún tipo de indicación concisa sobre la identidad de una materia al hombre⁶⁶⁴. Puede tratarse de una definición textual, descripción o nombre; puede tratarse de una representación visual o

⁶⁶³ PEPPER, S. *The TAO of Topic maps. Finding the way in the age of Infoglut*. En:

<http://www.gca.org/papers/xml europe2000/papers/s11-01.html>. Consultado el 18/02/2010.

⁶⁶⁴ PEPPER, S. ; SCHWAB, S. *Curing the web's identity crisis. Subject indicators for RDF*. En:

<http://www.ontopia.net/topicmaps/materials/identitycrisis.html>. Consultado el 15/02/2010.

sonora de una materia; o puede ser una combinación de éstas. Por tanto un indicador de materia es distinto de la materia a la que indica, ya que el indicador de materia sólo nos lleva a un documento que nos explica esa materia pero no es la materia en sí misma, sólo es una indicación de la misma.

Con un indicador de materia los usuarios podrían ser capaces de conocer exactamente a qué materia se les está referenciando. Siempre que las aplicaciones sean consideradas medios para la interacción hombre-máquina, los indicadores de materia proporcionarán una referencia común a los usuarios conectados a través de una aplicación, y de acuerdo con el indicador de materia podrá ser empleado tanto como una expresión denominativa del acuerdo como la identidad de una materia.

En tanto que los indicadores de materia son fuentes de información, se les tendrá que asignar siempre direcciones precisas. Estas direcciones podrán entonces ser empleadas por los ordenadores como una base para el establecimiento de la identidad, casi de la misma manera a como los ordenadores emplean las direcciones de materias direccionables. Si dos *topics* tienen el mismo indicador de materia, entonces por definición representan la misma materia y podrán ser procesados conjuntamente. Desde la perspectiva de una aplicación, si la identidad de dos *topics* es establecida por indicadores de materia que tienen la misma dirección, podrían ser considerados como representantes de la misma materia. A la dirección de un indicador de materia se le conoce como identificador de materia.

El indicador de materia y el identificador de materia son las dos caras del mismo mecanismo de identificación; el primero sería para el hombre y el segundo para las aplicaciones. Este mecanismo de identificación es la base para el acuerdo sobre la identidad de las materias de toda la red: entre aplicaciones, entre usuarios y entre aplicaciones y usuarios.

5.5.2.5. Materias Publicadas.

Las materias publicadas proporcionan un mecanismo para cualquier ordenador, así como para los usuarios en interacción con estos para poder saber cuando están tratando con un mismo documento. En otras palabras, definen la identidad de materias del discurso. En el discurso entre usuarios, la identidad se establece por un proceso basado en el uso de

nombres en contextos. El contexto es importante porque el mismo nombre puede ser empleado por diferentes materias, homonimia, y porque la misma materia puede tener más de un nombre, sinonimia. En ocasiones, el contexto es claro y la comunicación se establece inmediatamente; en otros casos, un complejo proceso de negociación tiene lugar hasta que se define un contexto común⁶⁶⁵.

Como sabemos, los ordenadores no pueden normalmente tratar con imprecisiones y necesitan contar con una identidad declarada de forma más precisa. En la web, la identidad se establece a través del empleo de URI's que proporcionan la dirección de la materia. Por lo tanto, "Recupera el artículo TAO de Pepper" lo traduce al siguiente envío [www.ontopia.net](http://www.ontopia.net/topicmaps/materials/tao.html): GET / *topic maps*/materials/tao.html HTTP/1.0.

Estos trabajan bien con materias que son fuentes de información, como en este caso, pero no del todo bien con otras materias. Por ejemplo, si queremos referirnos al autor del artículo antes mencionado, "Steve", ¿cómo lo hacemos? Podríamos hacerlo usando su dirección de email como una URI pero ¿qué ocurre entonces si queremos hablar acerca de la dirección de email de Steve? ¿No tendríamos que tener la misma URI, y no querría que resultase en Steve y su dirección e-mail, ambas materias bastante diferentes convirtiéndose en colaterales? El esquema de definición de materias que se ha descrito hasta ahora funciona bien en entornos controlados, pero es difícil sostenerlo en un contexto más amplio. Cualquier fuente de información puede ser considerado un indicador de materia simplemente por el hecho de estar recogido por una aplicación, sea o no deseada esa fuente por su editor para que sea un indicador de materia, y sea o no el editor consciente de su uso como un indicador de materia. De ahí que los indicadores de materia e identificadores definidos de este modo no sean fiables y puedan llegar a ser incluso ambiguos.

En cuanto a la ambigüedad, ¿cómo podemos saber exactamente qué materia es referenciada por el indicador de materia? Normalmente, los editores de documentos electrónicos no se plantean este problema porque no existe una intención específica, cuando se genera una fuente de información de proporcionar un indicador de materia para una única materia y por tanto no tiene sentido plantearse el grado de precisión necesario para evitar la ambigüedad. Además, cuando la fuente no ha sido publicada con la intención

⁶⁶⁵ PEPPER, S. *The case for published subjects*. En:

http://www.ontopia.net/topicmaps/materials/The_Case_for_Published_Subjects.pdf. Consultado el 12/02/2010.

expresa de ser empleada como un indicador de materia, no se puede esperar que ninguna organización sea la que establezca una estabilidad de la URI y la fuente de información a la que se refiere.

Si los editores fuesen conscientes de los defectos que se acaban de describir y quisieran proporcionar aplicaciones con indicadores e identificadores de materia desambiguados, la situación sería bastante mejor. Los editores pueden proporcionar colecciones de indicadores e identificadores de materias publicados de un modo normalizado, y declarar su intención de mantener su estabilidad, fiabilidad y disponibilidad en la red. Llegados a este punto, las aplicaciones llevarán implícitas materias publicadas, indicadores e identificadores.

El marco de publicación dentro del cual estas materias publicadas son empleadas es una red de aplicaciones interconectadas y de usuarios a los que se les permite acceder a esas aplicaciones. Puede ser desde un entorno tan amplio y abierto como la web hasta una red más o menos cerrada como la intranet de una empresa o el portal web de una comunidad o empresa. Es por esta razón que publicado no es sinónimo de público ya que su accesibilidad dependerá de la accesibilidad que se le proporcione a cada aplicación. Cualquier institución o usuario particular puede en principio publicar PSI's; sin embargo, no existe una autoridad que las procese y registre. Su adopción puede por lo tanto ser un proceso abierto y distribuido.

Los PSI's son normalmente publicados en colecciones que corresponden a un particular vocabulario controlado u ontología, o grupo de materias que son de especial interés para una organización, o aplicación, o individuo determinado. El hecho de que una colección particular de PSI's sea adoptada por otros dependerá de una serie de factores y no sólo de la visibilidad del editor y el grado de confianza de que disfrute⁶⁶⁶. La mayor autoridad de un editor es un factor destacado, pero el factor más importante es la probabilidad de que sus PSI's disfruten de un amplio consenso. No obstante, el concepto de autoridad varía según las circunstancias ya que se aplican distintos criterios según que el objetivo sea su adopción entre una comunidad relativamente acotada, tal como una institución jurídica, o si se trata de toda la comunidad web.

⁶⁶⁶ PEPPER, S. Op.cit.

Por tanto, la adopción de los PSI's es un proceso evolutivo en el cual sus colecciones estables y de autoridad van surgiendo y alcanzando una progresiva adopción. Sin embargo, el mecanismo es suficientemente flexible y abierto para que no haya necesidad de esperar al surgimiento de colecciones definitivas de PSI's. Cualquier usuario que necesite un PSI con un objetivo concreto debe considerar en primer lugar el empleo de uno que ya exista, y si no encuentra ninguno adecuado a sus propósitos entonces crear el suyo propio; si se crea un PSI más adecuado posteriormente, entonces sólo habrá que mapear el anterior sobre el actual PSI. Estos mapeos pueden crearse una sola vez y actualizarse repetidamente. Con el tiempo, a medida que vayan surgiendo nuevas normas, estos mapeos se hacen menos necesarios. Esto nos lleva a que no hay ningún motivo para empezar a usar inmediatamente materias publicadas, ya que los beneficios a corto y largo plazo en términos de interoperabilidad serán grandes dentro y entre aplicaciones.

Básicamente las recomendaciones⁶⁶⁷ que se deben seguir son:

1. Un identificador de materia publicado debe ser una URI.
2. Un identificador de materia publicado debe ser la respuesta a un indicador de materia publicado interpretable por el hombre.
3. Un indicador de materia publicado debe ser designado explícitamente por un único URI que debe ser empleado como su identificador de materia publicado.
4. Un indicador de materia publicado debe proporcionar metadatos comprensibles por el hombre acerca de sí mismo.
5. Un indicador de materia publicado debe proporcionar metadatos procesables por ordenador acerca de sí mismo.
6. Los metadatos definidos deben ser coherentes pero no tienen porqué ser equivalentes entre los comprensible por el hombre y los procesables por los ordenadores.

⁶⁶⁷ PEPPER, S. *Published Subjects: Introduction and Basic Requirements*. En:

<http://www.oasisopen.org/committees/download.php/2897/pubsub-pt1-1.01-cs.pdf>. Consultado el 18/02/2010.

7. Un indicador de materia publicado debe indicar qué elemento del documento es deseable que se emplee como PSI.
8. Un indicador de materia publicado debe ser capaz de identificar a su editor.

Como se dijo, los indicadores e identificadores de materia pueden ser declarados por cualquier usuario. Los URI's usados para identificar las clases y los elementos individuales que les pertenecen son, de hecho, identificadores de materia. En teoría nadie podría usar aquellos mismos identificadores de materia para señalar las mismas materias. Sin embargo, tenemos el problema de que nadie más sabe que estos identificadores de materia existen, e incluso si ellos lo hiciesen, no podrían estar seguros de su significado preciso porque las URI's no resuelven las fuentes de información interpretables, por la simple razón de que no ha habido una preocupación por crear indicadores de materia para ellos. Pero como se han creado identificadores de materia por parte de los usuarios para referirse a objetos digitales comúnmente aceptados, se ha hecho necesario publicar las URI's y crear indicadores de materia para ellos. Ese alguien, de mayor o menor autoridad, es quien decide los identificadores para un conjunto de materias y su publicación, en la confianza de que otros los usarán cuando necesiten referirse a las mismas. Cuando son publicadas de esta manera, con la voluntad de que sean usados por otros, se las llama "materias publicadas" y se habla de "colecciones PSI", donde el acrónimo "PSI" tiene dos extensiones: "indicador de materia publicada" e "identificador de materia publicada". Con ello, se refleja la dualidad de las materias publicadas: el aspecto humano interpretable, indicadores de materia, y el aspecto procesable del ordenador, identificadores de materia.

El empleo que se haga de un conjunto de PSI's dependerá de una serie de factores tales como su uso, las alternativas con que cuenten ya que cualquiera puede crear una colección de PSI y, más importante, si el editor es digno de confianza. La cuestión de la confianza es importante porque una de las principales razones para el empleo de PSI's es asegurar la interoperabilidad entre aplicaciones, incluso pensando en aquellas que puedan ir apareciendo. El valor de uso de los PSI's se incrementa casi proporcionalmente con su estabilidad y amplitud de adopción. La elección de una colección de PSI's que desaparece después de tres meses, o que nadie usa todavía porque hay una alternativa más estable, y los beneficios de interoperabilidad de lo que pudieron haber sido. Saber quién ha publicado un PSI es por lo tanto casi tan importante como conocer el PSI por sí mismo. Habiendo dicho

esto, es importante darse cuenta de que no se ha perdido todo si el PSI que uno ha escogido para usar se desactualiza ya que es posible crear un mapa desde un PSI en desuso a uno extendido. Para ilustrarlo consideremos el caso de los PSI's por países y lenguas. La especificación XTM 1.0 incluía una colección de materias publicadas para cada una de éstas, con PSI's tales como <http://www.topic-maps.org/xtm/1.0/language.xtm#en> para el inglés. Estos PSI's han sido ampliamente empleados en *topic maps*, si bien asegurando su adecuación con las recomendaciones desarrolladas por las materias publicadas de OASIS TC.

La principal entidad reguladora de los Published Subjects Indicators (PSIs) es el consorcio OASIS (Organización para el desarrollo de Normas de Información Estructurada⁶⁶⁸), consorcio internacional sin ánimo de lucro que dirige el desarrollo, convergencia y adopción de normas para comercio electrónico. Sus miembros elaboran una agenda técnica para procesos expresamente diseñados con el objetivo de promocionar la interoperabilidad de lenguajes de marcado. Editan normas internacionales para servicios Web, esquemas XML, publicaciones electrónicas, *topic maps* e interoperabilidad dentro y entre mercados. Fundado en el año 1993 bajo el nombre de SGML Open, como un consorcio de vendedores y usuarios dedicados al desarrollo de guías para la interoperabilidad entre productos basados en el Lenguaje de Marcado General Normativo (SGML). OASIS cambió su nombre en 1998 para reflejar su aplicación a un ámbito más extendido de los lenguajes de marcado, de forma que pudiera comprender el Lenguaje de Marcado Extensible (XML) y otras normas relacionadas.

Dentro de OASIS se ha creado un comité técnico denominado “Materias Publicadas para *Topic Maps*”. El objetivo de este comité es promocionar el empleo de materias publicadas mediante la elaboración y difusión de recomendaciones, requisitos para su definición, organización y aplicación. Las materias publicadas son definidas conceptualmente en la norma ISO 13250 de *topic maps* y han sido más detalladas en la especificación 2.0 de la sintaxis XTM, XML para *topic maps*.

Estas materias publicadas verifican lo que ninguna URL o URN ha sido capaz hasta ahora de hacer en términos de declaración de identidad de un modo progresivo y coherente. Mientras que las URL's pueden sólo declarar la identidad de las fuentes de información, los

⁶⁶⁸ OASIS. En: <http://www.oasis-open.org/who/>. Consultado el 18/02/2010.

PSI's pueden establecer la identidad de cualquier recurso digital. Mientras que los URN's son estrictamente controlados por el IETF (Internet Engineering Task Force) y no tienen un fácil mecanismo de asignación y por lo tanto son difíciles de interpretar, los PSI's pueden ser definidos por cualquiera y resolver de forma sencilla la asignación de indicadores de materia interpretables por el usuario.

Según Pepper, existe un proceso evolutivo hacia una mejora de la confianza y usabilidad de colecciones de PSI's para la designación de materias que el hombre y las máquinas necesitan interpretar. Esto supondrá importantes beneficios en términos de interoperabilidad para el desarrollo de las bibliotecas digitales semánticas.

5.6. Principios generales para la edición de un *Topic map*.

La norma ISO/IEC 13250 define los conceptos y sintaxis para el intercambio de *topic maps* entre diferentes motores de edición *topic map*. Sin embargo, no define cómo es posible crear y mantener un *topic map*. Esta tarea se ha abandonado tradicionalmente a los usuarios y las orientaciones que los motores *topic map* puedan proporcionar.

Los *topic maps* pueden ser editados de muchas maneras, incluyendo su edición manual con sintaxis XTM, o con sintaxis LTM, o bien la conversión de datos fuente estructurados de las bases de datos, documentos XML, o sencillamente empleando una herramienta de edición y diseño de *topic maps*. En cualquier caso, empezar desde los datos fuente proporciona la posibilidad de asumir un compromiso flexible de calidad con el mismo, además de facilitar posteriores trabajos de modificación y enriquecimiento.

En general, podemos decir que existen tres vías para la creación de *topic maps*: i) manualmente, ii) automáticamente, iii) semiautomáticamente. En cualquier caso, existen dos niveles diferentes en todo *topic map*: a) el nivel de las clases, b) el nivel de las instancias. Donde las clases se crean en la fase de diseño de la ontología *topic map*, en tanto que las instancias son creadas en la fase de edición del *topic map*; en cualquier caso, cada fase emplea sus propias técnicas.

Nosotros vamos a centrarnos, por existir hasta el momento más ejemplos al respecto, en la vía de generación manual de *topic maps*. Para lo cual, a modo de guía, indicaremos las

principales etapas en la generación de un *topic map*. Esto nos permitirá más adelante valorar las posibilidades de generación automática de *topic maps*.

Lo primero que un autor de *topic maps* ha de plantearse son los objetivos que persigue alcanzar con la ontología del *topic map*. Para ello, se hace necesaria la consecución de las siguientes tareas clave⁶⁶⁹:

1. Definición del dominio de aplicación. Lo que debe ser cubierto por parte del *topic map*.
2. Definición de los requisitos funcionales tales como quién empleará el *topic map* y con qué propósitos.
3. Definición de un Schema que concrete los tipos de materias a cubrir y cómo éstas deben relacionarse entre sí para que *topic map* sea coherente.
4. Selección de herramientas y aplicaciones. La arquitectura de información y el software que debemos usar para poder ejecutar la aplicación.
5. Poblamiento del *topic map*. La generación manual o automática de instancias, siempre y cuando éstas sean capaces de satisfacer las necesidades de los usuarios a quienes va destinado este *topic map*. También debe evaluarse la adecuación del schema al tipo de instancias que están siendo introducidas.
6. Mantenimiento tanto del *topic map* como de la aplicación con la que se ha editado. Para ello es aconsejable utilizar regularmente la aplicación para saber si hay actualizaciones, además de añadir o modificar los posibles elementos que los usuarios demanden al *topic map* durante el proceso de navegación por los recursos digitales de la biblioteca digital semántica.

Una vez planteados estos pasos, que en general se deben seguir para el diseño de un *topic map*, nos planteamos cómo diseñar la ontología que rige el comportamiento del *topic map* a editar. Para ello, debemos tener en cuenta que trabajamos en este nivel con clases o *topic*

⁶⁶⁹ RATH, H. *The Topic maps handbook*. Germany : Empolis, 2003. p.27. En: [http://www.sts.tu-harburg.de/~r.f.moeller/lectures/anatomie-i-und-k-system/empolistopic mapswhitepaper_eng.pdf](http://www.sts.tu-harburg.de/~r.f.moeller/lectures/anatomie-i-und-k-system/empolistopic%20mapswhitepaper_eng.pdf). Consultado el 15/12/2009.

types, que al igual que las materias son un conjunto de propiedades en común. Sin embargo, no es posible el diseño si no somos capaces de relacionar las clases entre sí, para lo que hace falta generar un conjunto de normas o gramática que permita la validación durante la edición del *topic map*⁶⁷⁰. A estas reglas se les denomina “constraint”. Son un poderoso mecanismo tanto para el control del proceso de edición como de validación semántica del contenido del *topic map*. Para la normalización en la edición de estas reglas contamos con el ya visto TMCL.

En el proceso de diseño, se definen en TMCL las reglas a emplear para los distintos elementos del *topic map*. Estas definiciones pueden venir ya dadas en el propio motor *topic map*, o bien ser indicadas por el propio diseñador para actualizaciones futuras. Por ejemplo, algunas reglas típicas son:

Reglas para la clase *topic* “libro”: Todas las obras publicadas sobre técnica jurídica posteriores a 1977.

Nombres: mínimo un título en el ámbito de la lengua española.

Occurrences: “tejuelo” exactamente uno, “páginas” al menos 25.

Role que juega. Participa como “ítem de la colección” en asociación con el empleo del *role* “colección continente”.

Reglas para la clase *association* “ítem de la colección”.

Roles: exactamente “un ítem de la colección”, y exactamente una “colección continente”.

Clases *topic* de las funciones *role* para los *roles* “ítem de la colección” y “colección continente”. “colección”----“libro”.

⁶⁷⁰ RATH, H. ; MOORE, G. *Topic map Constraint Language (TMCL) Requirements and Use Cases*. En: <http://www.isotopicmaps.org/tmcl/requirements.html>. Consultado el 15/12/2009.

La declaración de una clase con sus reglas permite la agrupación de instancias en lo que se conoce por *template*. Una colección de templates define una clase de *topic map*⁶⁷¹.

Una vez hemos dejado claro los objetivos a cubrir por nuestro *topic map*, así como el punto de partida, nos planteamos el comienzo de la edición de la ontología *topic map* en las siguientes fases:

1. Fase de análisis.
2. Fase de diseño.
3. Fase de creación del *topic map*.
4. Fase de evaluación.
5. Fase de documentación.

1. Fase de análisis⁶⁷².

En esta fase nos aseguramos de que la ontología *topic map* va a poder generar un *topic map*. Esto obliga a un trabajo previo de: i) enumeración de los requisitos que la ontología *topic map* debe de cubrir y asignar sus prioridades. Por ello, es tan necesaria la consulta a los usuarios sobre sus necesidades y demandas más habituales, las necesidades de diseño; así como la consulta a los servicios técnicos del centro para conocer las posibilidades tanto técnicas como de personal y presupuesto con que cuenta para crear y mantener un *topic map*. ii) Enumeración de los lenguajes documentales con los que cuenta nuestro centro y que pudieran ser reutilizados por el *topic map*. iii) Representación en un mapa conceptual de todas las materias del dominio de aplicación.

i) Los requisitos influyen en las decisiones futuras de diseño que se tomen, especialmente aquellas relacionadas con los requisitos de los usuarios. Es por ello que hemos recogido un

⁶⁷¹ RATH, H. Topic map Fundamentals for Knowledge Representation. En PARK, J. XML Topic maps. Boston : Addison-Wesley, 2002. pp.360-361.

⁶⁷² OBRST, L. ; LIU, H. Knowledge Representation, Ontological Engineering, and Topic maps. En PARK, J. XML Topic maps. Boston : Addison-Wesley, 2002. pp.126-129.

conjunto de preguntas de las que tanto los diseñadores de la ontología *topic map*, como los especialistas en el dominio del derecho, el personal técnico de la biblioteca y los usuarios pueden servirse como orientación⁶⁷³. Sirven como guía en la fase de análisis y ayudan a estructurar los requisitos en categorías: requisitos del sistema, de los usuarios, del diseño. Todos ellos tienen asignada una prioridad o un peso para ponderar la prioridad de cada uno de ellos.

Rath, nos indica la existencia de cuatro indicadores para la agrupación de las posibles preguntas que realizar a los responsables de una biblioteca para el diseño del *topic map*⁶⁷⁴. Estos son:

I) Indicador acerca del diseño general del *topic map*. En nuestro caso para el *topic map* de la biblioteca digital semántica del CEPC:

¿Dónde están los límites del dominio de la aplicación? La pregunta es muy importante dado que en un *topic map* la exhaustividad y precisión del conocimiento recogido es ilimitado, pero debe ser limitado para que pueda ser gestionado. Los límites definidos afectan a la profundidad de la materia, precisión, y a su amplitud, exhaustividad, lo cual redundará en el diseño del *topic map* y su poblamiento. Además, permite dilucidar en qué dirección cambiará el dominio en el futuro.

¿Quiénes son los usuarios y cuáles son sus expectativas? ¿Existen grupos de usuarios con necesidades específicas? El *topic map* tiene que tratar de satisfacer las necesidades y expectativas de su comunidad. Esta debe ser tomada en cuenta en cada una de las etapas del diseño. Si el *topic map* está orientado a distintos perfiles de usuario, entonces habrá que ponderar el peso de cada perfil de usuario a la hora de diseñar el *topic map*.

¿Cuáles son los límites económicos y de tiempo? ¿Con qué recursos humanos contamos? El principio del *topic map* permite su aplicación a un sistema de información a un coste muy bajo y su posterior mejora. Por ello, es importante que al comienzo el *topic map* no entre en

⁶⁷³ Veremos algunas de las principales cuestiones que realizamos a los responsables de la biblioteca del CEPC y explicaremos su interés con ejemplos. En el siguiente capítulo se mostrarán las preguntas agrupadas por indicadores con sus correspondientes ponderaciones en tablas.

⁶⁷⁴ RATH, H. *The topic maps handbook*. Germany : Empolis, 2003. p.29. En: <http://www.uni-giessen.de/germanistik/ascl/gldv2001/proceedings/pdf/GLDV2001-rath.pdf>. Consultado el 15/12/2009.

muchos detalles, lo que lo convertiría en caro y puede que incluso poco operativo. El diseño inicial debe adecuarse a las principales necesidades del usuario.

¿Debe la topología del *topic map* estar restringida? Un *topic map* con una arquitectura en árbol es lo suficientemente exhaustiva para la navegación, como para tener que acudir a diseños multidimensionales con muchas *associations*. Es aconsejable la restricción del número de jerarquías en el diseño en árbol para evitar una clasificación demasiado detallada.

¿Se desea mostrar el *topic map* analíticamente o con un modelo de representación más visual? Algunas aplicaciones *topic map* muestran todos los detalles del *topic map* al usuario, con sus clases, *topics*, nombres, *occurrences*, *associations* y scopes. Otras aplicaciones ocultan los detalles del *topic map* y proporcionan una representación abstracta del mismo al usuario presentándole únicamente aquellos objetos del interfaz con los que se encuentra familiarizado. La primera categoría de aplicaciones sigue el modelo de las taxonomías y tesauros. La segunda categoría es acerca de *topic maps* que utilizan metadatos en la gestión documental y la gestión de contenidos ocultos por los interfaces de usuario de las aplicaciones.

II) Indicador acerca del uso del scope en el diseño del *topic map*.

¿Son necesarios distintos puntos de vista? El scope o ámbito permite la presentación de distintos puntos de vista sobre el *topic map* y proporciona la capacidad de filtrar distintas partes del mismo. Las vistas permiten modelar el acceso a la información por parte del usuario en función de sus intereses, las lenguas o los derechos de acceso.

¿Es necesaria la edición de un *topic map* multilingüe? Los nombres multilingües de los *topics* se manejan por medio del scope; sin embargo, el consorcio OASIS proporciona una colección de materias publicadas para muchas lenguas que pueden ser utilizadas para los *topic maps* multilingües, junto con los tesauros y redes léxicas multilingües como Eurowordnet.

III) Indicador sobre el poblamiento del *topic map*.

¿Cómo será el poblamiento del *topic map*? Un *topic map* puede ser editado manualmente, automáticamente o bien semiautomáticamente. La calidad esperada del resultado es el criterio a considerar para decidir por cuál de los modos nos decantamos. Normalmente un *topic map* manual es más preciso que el semiautomático.

¿Es necesario tratar con información legada? Nos referimos por información legada a la ya existente en otros lenguajes documentales, la cual puede ser migrada o convertida en un *topic map*. La información legada existe en instituciones y organizaciones que ya cuentan con sus propios lenguajes para la recuperación de información en las bases de datos de su dominio de actividad. Así por ejemplo, un punto de partida que empleamos para el poblamiento del *topic map* fue el tesoro EUROVOC, por tratarse de un tesoro de derecho multilingüe. Sin embargo, en ocasiones es difícil encontrar recursos apropiados, EUROVOC es un tesoro de derecho comunitario pero no de historia del derecho, e integrarlos en nuestro *topic map*.

¿Qué recursos de información son susceptibles de convertir a *occurrences*? La identificación de aquellos recursos susceptibles de ser convertidos a *occurrences* en la fase de diseño obliga a contar no sólo con el personal de biblioteca, sino sobre todo con especialistas en la materia tratada por estos recursos. En general, debemos partir de que es mejor limitar los recursos para obtener una cantidad. O bien, categorizarlos y tratar únicamente con categorías a fin de no tener que tratar con todos los recursos en la fase de diseño.

¿Hasta qué punto debe estar el *topic map* ligado con los recursos de información? La creación manual y mantenimiento de *occurrences* puede ser una tarea costosa, por tanto se debe tener en cuenta la creación y mantenimiento automático de las *occurrences* por medio de una herramienta de clasificación o categorización.

IV) Indicador de temas editoriales.

¿Quiénes son los autores y cuáles sus ámbitos de especialidad? Para asegurarse que los autores del *topic map* tiene conocimiento sobre el dominio y las necesidades de los usuarios.

¿Cuántos autores trabajarán con el *topic map*? el *topic map* creado por un único autor estará más limitado en sus posibilidades que el creados por varios especialistas. Esta pregunta

tiene varias implicaciones en el sistema y el trabajo de edición. Se debe tener en cuenta que el schema *topic map* no puede ser cambiado cuando se esté poblando el *topic map*. Únicamente se debe cambiar el schema cuando ningún autor haya incluido *topics*.

¿Cómo tener en cuenta los derechos de acceso y la gestión del usuario? Distintos grupos de autores pueden tener distintas competencias en la edición del *topic map*. Las acciones permitidas al usuario dependen de la clase de *topic* y el grupo de usuarios en el que se centre cada autor.

¿Es necesario distinguir entre el *topic map* continuamente editado y aquel con el que los usuarios operan? Un peldaño en la publicación puede separar estos dos *topic maps*. El *topic map* autorizado es copiado periódicamente y la copia llega a ser el *topic map* del usuario. La copia puede ser completa o una parte del *topic map*. La parte puede ser identificada por el scope o por una consulta *topic map*.

¿Cómo será tratado el mantenimiento del *topic map*? Asegura que la información desactualizada o invalida es borrada a intervalos regulares; es decir, se planifica cuando y cómo se realizan las actualizaciones automáticas.

¿Será la fusión de *topics* un problema en el futuro? Investigar si las colecciones de materias publicadas están disponibles o si tienen que ser definidas. Formar a los autores en la asignación de las materias publicadas a los *topics* siempre que sea posible, además de sugerir cómo tratar con las reglas de denominación de los *topics*.

¿Es importante poner atención en la ratio entre *topics*, *occurrences* y *associations*? Pues sí, dado que la comprensión del *topic map* depende de estas ratios. Por ejemplo: un *topic* asociado con miles de otros *topics* dificulta la navegación. No obstante, puede que el *topic map* no esté destinado a ser navegable o visualizable sino a ser usado con otros propósitos como servir a modo de repositorio.

ii) Aprovechamiento de otros lenguajes documentales.

Las bibliotecas, y en nuestro caso la biblioteca del CEPC, cuenta para sus labores de tratamiento documental de la colección con un conjunto de lenguajes propios y externos

tales como el tesoro de derecho del Ministerio de Presidencia, el tesoro EUROVOC revisado por la biblioteca del Senado, o la lista de encabezamientos de bibliotecas públicas. Según Garshol⁶⁷⁵, los *topic maps* deben partir en su diseño de los recursos y repositorios existentes en el centro, o incluso ser utilizados para la migración de su información a un *topic map*. Esto obliga a una labor previa de evaluación de estos recursos para saber si pueden ser utilizados, de qué modo, y qué modificaciones deben ser efectuadas. Asimismo, aconseja la localización de lenguajes documentales especializados o de los dominios para los que se va a editar el *topic map*. Entre ellos podemos contar con: ontologías, esquemas de metadatos, sumarios, índices, taxonomías, DTD's y schemas XML, o incluso con tesauros contruidos con *topic maps* como el construido por Pepper y que se muestra en la figura siguiente⁶⁷⁶.

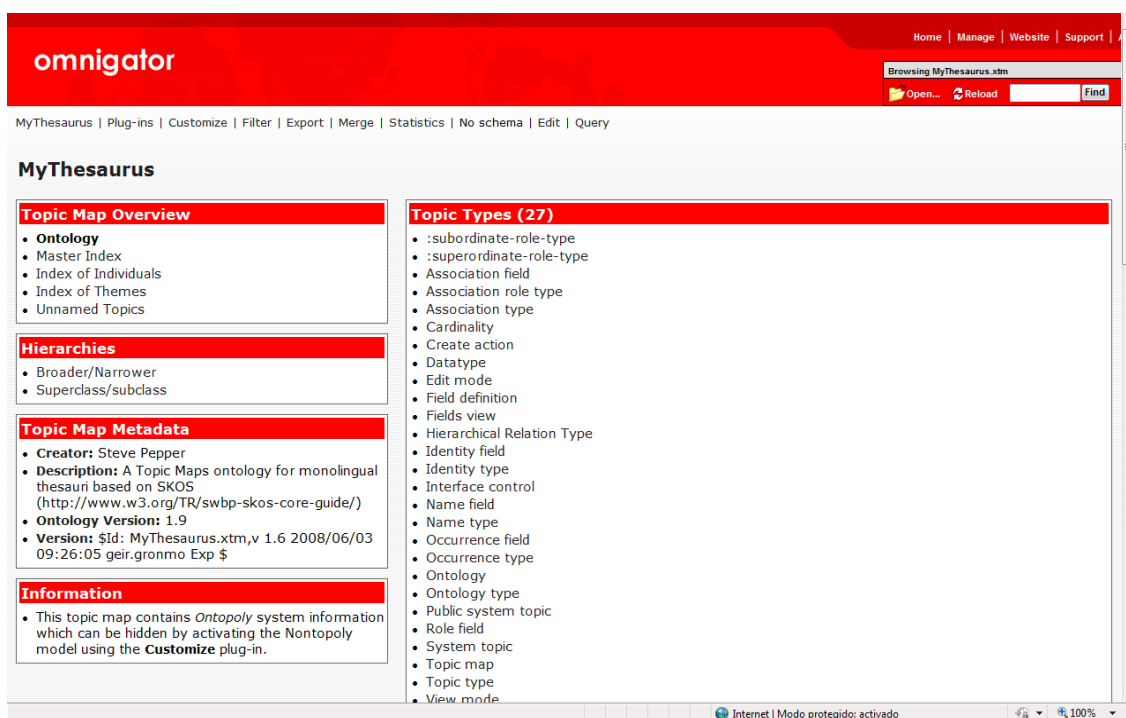


Figura 52. Ejemplo de tesoro editado con Omnigator por Pepper.

iii) Representación en un mapa conceptual de todas las materias del dominio de aplicación.

⁶⁷⁵ GARSHOL, L. Towards a methodology for developing topic maps ontologies. En MAICHER, L. SIGEL, A. ; GARSHOL, L. *Leveraging the semantics of topic maps*. Berlin : Springer, 2007, p.25-26.

⁶⁷⁶ PEPPER, S. Mythesaurus. En: http://localhost:8080/omnigator/models/topic_map_complete.jsp?tm=MyThesaurus.xtm.

Consultado el 15/01/2010.

Los *topics* trasladan las designaciones de las materias de los lenguajes documentales de un dominio de conocimiento a un lenguaje procesable por una aplicación, en nuestro caso una aplicación *topic map*. Para ello el punto de partida es hacer una lluvia de materias o “brainstorm”, que tengan que ver con nuestro dominio, y representarlas por medio de una aplicación mapa conceptual⁶⁷⁷. Con ello, se puede enumerar cualquier objeto con independencia de la clase de *topic* a la que vaya a ser traducido en nuestro *topic map*: como una clase *topic*, un *topic*, una *association topic* o una *occurrence topic*. Esto nos permite, desde el primer momento del diseño del *topic map*, considerar los distintos enfoques que deseamos presentar a nuestros usuarios para su consulta. Y aunque se trata de un proceso desorganizado, dota al diseñador de creatividad en el dibujo de un primer marco de estructuración de las materias para su traslación a *topics* procesables por una aplicación *topic map*. Esta primera lista estructurada de materias de la que se parte en la fase de diseño tiene por funciones⁶⁷⁸: a) Asegurarse de que todas sus materias sean únicas, evitando la sinonimia entre las mismas. Para ello, cuando se encuentran en el mapa conceptual materias sinónimas se guardan como denominaciones alternativas para una única materia. b) Ver las materias listadas e identificar las clases para los *topics*, las *occurrences* y las *associations*; así como indicar cuáles pueden ser los *associations* roles. Asimismo, es posible identificar las instancias de *topics*, *occurrences* o *associations*. c) Enumerar las materias que pueden ser utilizadas de forma agrupada en cada ámbito para la definición de los distintos contextos informativos.

⁶⁷⁷ OBRST, L. ; LIU, H. Op. cit., p.127.

⁶⁷⁸ RATH, H. *The topic maps handbook*. Germany : Empolis, 2003. pp.30-31. En: <http://www.uni-giessen.de/germanistik/ascl/gldv2001/proceedings/pdf/GLDV2001-rath.pdf>. Consultado el 15/12/2009.

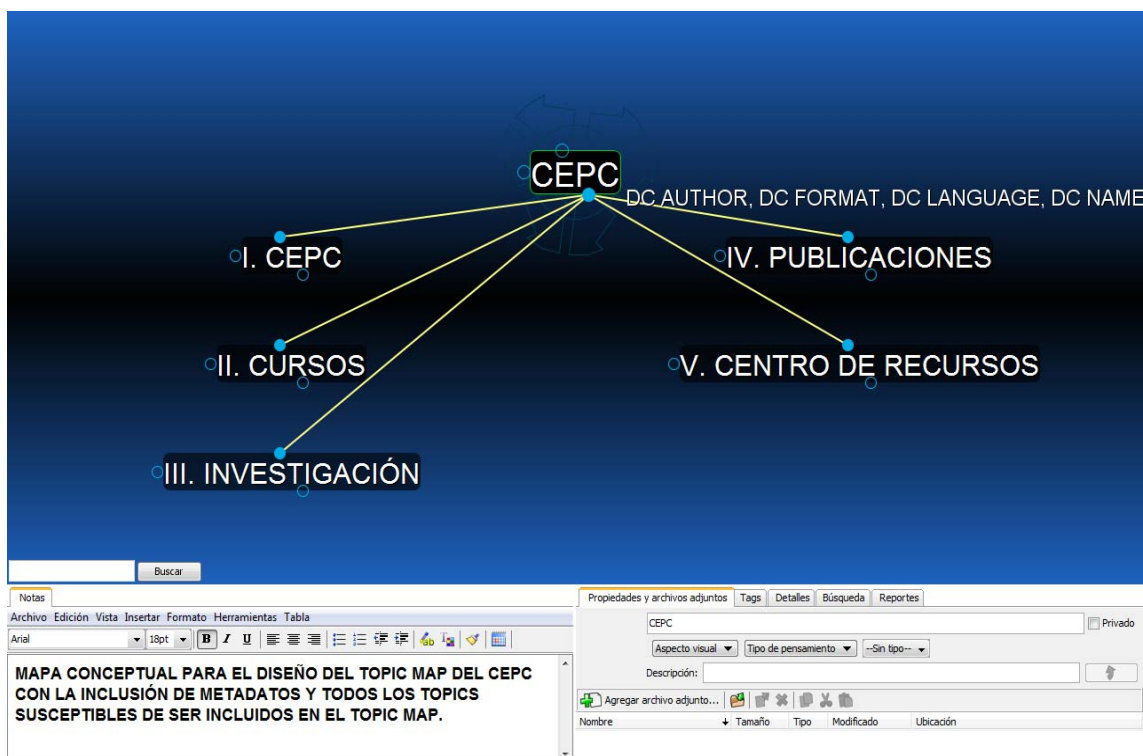


Figura 53. Mapa conceptual editado con la herramienta TheBrain para el diseño de la ontología *topic map* del CEPC.

2. Fase de diseño.

A partir del mapa conceptual ya se cuenta con una primera estructura de clases y *associations* roles seleccionados para comenzar la fase de diseño del *topic map*. En esta fase se pasa a organizar y definir las clases de manera más detallada. Esto quiere decir que cada clase debe ser definida como una plantilla en la que se incluyan los *topic* y sus reglas. Será esta colección de plantillas la que definirá el schema del *topic map*, siempre contando con las reglas de la norma para TMCL para lograr un diseño más claro y mejor documentado del *topic map*. En cualquier caso, Rath nos indica que en esta fase, la cuestión clave es la identificación clara y precisa de las reglas que operan para este *topic map*⁶⁷⁹.

Para poder expresar las reglas que rigen la inclusión de *topics*, se debe comenzar por escribir las reglas de inferencia que deseamos se cumplan, al igual que se hace en el diseño de Chen para el diseño de modelos relacionales en bases de datos. Para ello, nos podemos ayudar de un lenguajes de diseño de software como UML o bien, como veremos en el siguiente capítulo, por medio de algún programa específico para la edición de reglas con schemas en

⁶⁷⁹ RATH, H. Op. cit.

TMCL como ONOTOA⁶⁸⁰. Programa éste que presenta una interfaz que simula la programación con sintaxis UML de reglas de inferencia para schemas *Topic map*. De hecho, para el diseño de nuestro *topic map* hemos contado con el trabajo de Niederhausen⁶⁸¹.

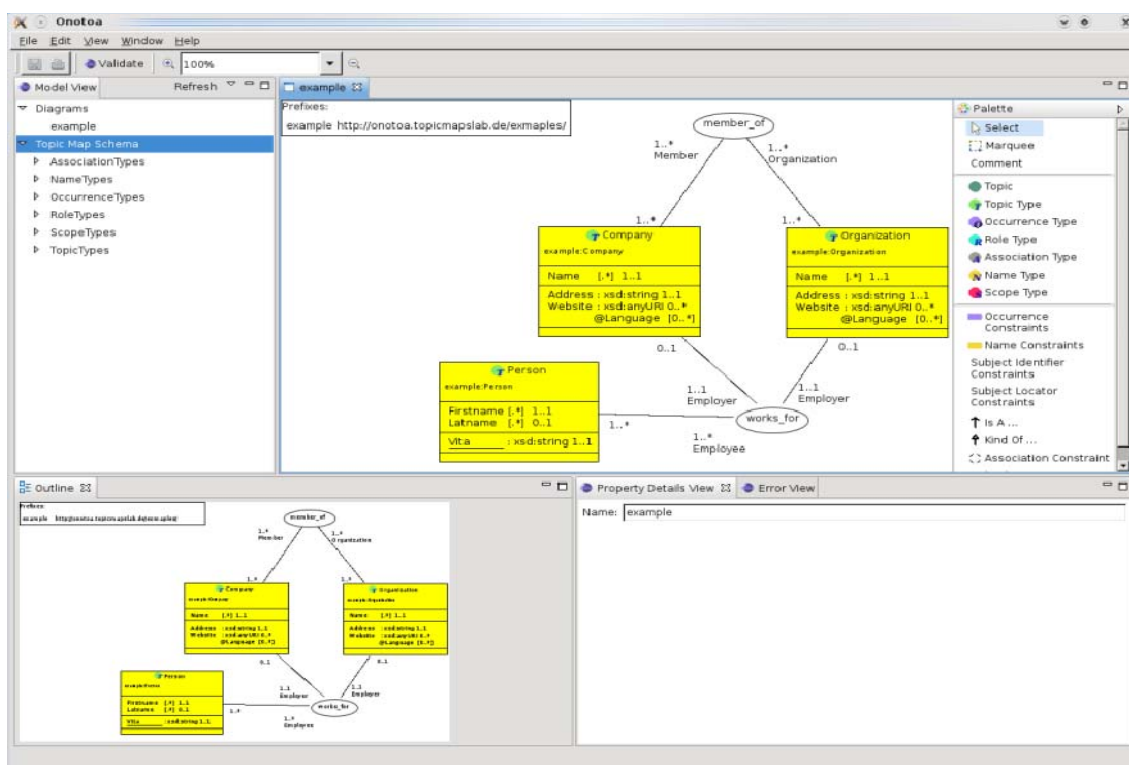


Figura 54. Ejemplo de diseño de un schema *topic map* con el programa ONOTOA.

Niederhausen y Rath nos indican algunas reglas para la definición de las clases en cada uno de los elementos esenciales del *topic map*, su TAO⁶⁸². Vamos a ver cuáles son éstas:

- **Para la definición de las clases *Topic*.** Se debe distinguir entre recurso, instancia *topic* y clase *topic*. Al respecto, la norma es que sólo los recursos cosificados como *topics* pueden tener propiedades: un nombre, una *occurrence*, o estar asociado a otros *topics*. Si un recurso no necesita contar con nombres, *occurrences*, o estar asociado entonces es que el recurso es únicamente una *occurrence*, pero no un *topic*.

⁶⁸⁰ ONOTOA. En: <http://onotoa.topic-mapslab.de/>. Consultado el 15/12/2009. Se trata de un software diseñado por Niederhausen para su tesis doctoral, fruto de su trabajo en el grupo *Topic map Labs* dirigido por Lutz Maitcher de la Universidad de Leipzig.

⁶⁸¹ NIEDERHAUSEN, H. *Creating a topic map ontology with Onotoa*. En: <http://onotoa.topic-mapslab.de/tutorial>. Consultado el 15/12/2009.

⁶⁸² Id.

La plantilla *topic* permite limitar las posibles propiedades de las instancias, tales como: el número de nombres en un ámbito; las clases de *occurrences* en un ámbito así como su cardinalidad; la distinción entre reglas optativas, obligatorias y las no permitidas; así como las *association* roles que las instancias juegan en cada una de las clases.

- **Para la definición de las clases *Occurrence*.** Se comprueba la colección documental con la que se cuenta, así como el lenguaje de especialidad del dominio en el que se encuentra para la identificación de las clases de *occurrences*. Observando cada clase de *topic* y *association* se puede averiguar cuáles son las principales clases de *occurrence*.

La plantilla *occurrence* permite limitar el empleo de instancias en clases *occurrence* mediante: i) la asignación de las clases *occurrence* a las correspondientes instancias de clase *topic*, ii) el ámbito, iii) el número de nombres en un ámbito y, iv) para la distinción entre restricciones optativas, obligatorias y no obligatorias.

- **Para la definición de las clases *association*.** Rath nos aconseja ver todas las clases *topic* y sus posibles combinaciones para averiguar las clases *association* significativas⁶⁸³. Las clases *topic* detalladas pueden en ocasiones ser expresadas como una clase *topic* general con una *association*. Por ejemplo: el CEPC es una instancia de la clase “organismos consultivos del Estado”, pero también puede entenderse como que el CEPC es una instancia de organismos y que el CEPC está proporcionando “informes”

En general, durante el diseño del *topic map*, debemos tener presente que cuando se introducen clases *topic* y clases *association*, la denominación de las clases *association* deben realizarse por medio de un sustantivo que exprese la relación como puede ser el lugar de la sede del CEPC.

La denominación del *association* role debe hacerse con un sustantivo⁶⁸⁴, que puede ser una subclase o superclase de esa clase *topic* en la que actúan los roles. Por ejemplo, una superclase de CEPC es institución y la superclase de la localización puede ser país.

⁶⁸³ RATH, H. Op. cit., p.32.

⁶⁸⁴ Ibid., p.33.

La denominación del enlace o *arc*, en el caso de relaciones binarias debe ser un verbo que exprese el significado cuando consiga de un role que actúe como *topic* de otro⁶⁸⁵. Por ejemplo: El CEPC está localizado en Madrid/ Madrid es la localización del CEPC, “está localizado” es el enlace entre CEPC y Madrid y se logra que el *role* localización, Madrid, sea un *topic*.

Finalmente, debemos tener en cuenta que la plantilla *association* restringe el empleo de instancias de la clase *association* por diversos motivos⁶⁸⁶: i) tanto por el ámbito como por los roles *association* con sus cardinalidades, ii) por la combinación de clases *topic* a las que se permite actuar como roles, iii) por la forma de etiquetar los arcs o enlaces, iv) por el número de nombres en el ámbito, v) por las clases *occurrence* de cada ámbito y su cardinalidad, vi) por los roles *association* que las instancias de cada clase *association* juega en la asociación de un clase determinada, o bien por vii) la distinción entre las reglas obligatorias, optativas y las imposibles.

Así por ejemplo, el Centro de Estudios Políticos y Constitucionales o CEPC. Las reglas de una de sus clases *association* pueden ser contextualizadas por “historia del derecho”, debe contar con el role “organismo” y otro role “localización”, las clases *topic* permitidas para los roles antes mencionados pueden ser “centro de estudios” o “centro consultivo”.

Junto a la definición de las reglas para definir los constituyentes esenciales del *topic map*, resulta bastante útil la consideración de reglas para:

- la definición de los contextos. El contexto expresa la validez de las características *topic*, *occurrence* y *association*. Es empleado para definir las distintas vistas en el *topic map*. Asimismo, los contextos tienen un papel clave en la unión de los *topic maps* por nombre, proporcionando sinónimos de los mismos. Analiza las vistas necesarias y fusiona los sinónimos para la identificación de los *topics* propios de cada contexto. Estos *topics* contextualizados pueden ser instancias de clases empleadas para la agrupación de *topics* y que no estén aislados.

⁶⁸⁵ Ibid., p.33.

⁶⁸⁶ AHMED, K. [et al.]. *Professional XML Metadata*. Birmingham : Wrox Press, 2001. pp.258ss.

-la declaración de jerarquías de clases. Las jerarquías de clases constituyen una parte esencial de cualquier tipo de ontología. Modelan las taxonomías y son la base de la coherencia, la inferencia y el potencial de búsqueda de un *topic map*. Sin embargo, para el modelado de una jerarquía de clases se hace necesaria la definición de una clase *association* en caso de que las clases no sean tales desde el punto de vista de los *topic maps* porque la *association* superclase-subclase no pueda ser aplicable. Esto se produce fundamentalmente en *topic maps* con múltiples jerarquías de clase. En nuestro caso, para el CEPC, contamos con múltiples jerarquías como en publicaciones, cursos, etc., que impiden establecer una relación de superclase-subclase entre todos los *topics*.

-la declaración de identidades de materia. Si queremos preparar el *topic map* para una fusión correcta con otros, debemos establecer la identidad de materia de los topics. Esto tiene que ver tanto con el uso de las materias direccionables del dominio de aplicación como con el empleo de identificadores de materia para aquellas que no son direccionables. Puede ocurrir que las materias publicadas estén disponibles para un dominio, y sus identificadores de materia puedan ser usados directamente. Cuando es necesario tanto definir un identificador de materia como los indicadores de materia, es muy útil partir de los vocabularios controlados existentes sobre la materia. En nuestro caso, tuvimos que partir del tesoro de derecho del CINDOC, del tesoro del CEPC y del EUROVOC.

3. Fase de creación del *topic map*.

Hasta ahora el diseño se ha realizado sin crear ningún *topic map* real sino que ha sido un borrador de trabajo. En esta fase explicamos cómo transformar el diseño en un *topic map* real. Para ello, existen tres posibles técnicas⁶⁸⁷:

Por medio de la escritura del *topic map* en sintaxis XTM con un editor de XML.

Con la escritura con una sintaxis susceptible de ser transformada automáticamente en XTM.

O bien, empleando una herramienta de edición de *topic maps* que proporcione un interfaz de usuario para su edición.

⁶⁸⁷ GOLDFARB, CH. ; PRESCOD, P. *XML handbook*. Upper Saddle River : Prentice hall, 2002. pp.643ss.

La escritura con sintaxis XTM ocasiona errores cuando se deben introducir muchos *topics* e interrelacionarlos entre sí manualmente con la sintaxis de las *association*. El empleo de sintaxis alternativas como LTM o ZTM tiene el mismo problema, si bien se reducen los errores porque no hace falta escribir tanto código. En cambio, con un interfaz de edición automática se oculta la complejidad de la sintaxis, el tratamiento y la interrelación entre *topics*. Estos programas hoy en día son escasos, especialmente cuando se trata de software libre.

4. Fase de evaluación.

La evaluación, especialmente cuando la realizamos contando con especialistas en el área de especialidad a la que se orienta el *topic map* es clave para que resulte útil a la comunidad a la que va orientado. El diseñador, desde el comienzo plantea cuestionarios a los técnicos del centro para evaluar cómo desea resolver las dudas informativas a su comunidad de usuarios. Y a éstos, se les plantean una serie de preguntas para conocer sus capacidades de búsqueda y habilidades para la resolución de sus demandas informativas con los recursos que el centro pone a su disposición⁶⁸⁸. Pero es que además, una vez creados los pertinentes *topics*, *occurrences*, y *associations* con su adecuada colección de contextos a los que se adaptan las características de los *topics*, el diseñador navega por el *topic map* comprobando la satisfacción de una serie de preguntas tales como: ¿es coherente el *topic map*?, ¿la granularidad o nivel de especificidad de los *topics* es la adecuada, así como cuántas subclases debería tener cada una de las clases?, ¿las restricciones son las adecuadas?, ¿los *names* son los adecuados?

Una vez el diseñador ha terminado las comprobaciones y mejorado el *topic map* de acuerdo con las respuestas a estas preguntas, se debe incluir a los distintos tipos de usuario en la evaluación, mientras que el diseñador se plantea la satisfacción del usuario acerca de la navegación por el *topic map*⁶⁸⁹. Junto a la evaluación realizada por parte del usuario se pueden efectuar otros tipos de pruebas evaluadoras tales como las pruebas de consistencia de las normas impuestas al *topic map* y recogidas en la plantilla del *topic map constraint language*. Otra prueba es la evaluación estadística para averiguar cuál es el número de *topics*,

⁶⁸⁸ YI, M. Information organization and retrieval using a topic maps-based ontology: Results of a task-based evaluation. *Journal of the American Society for Information Science and Technology*. Vol. 59, nº12. pp.1898-1910.

⁶⁸⁹ Ibid., p.1908.

associations y *occurrences* en cada una de las clases específicas, o cuántos roles y cuáles realizan los *topics* en las *associations*. Finalmente, planteamos la evaluación funcional para conocer si los objetivos funcionales que nos planteamos en la fase de diseño son satisfechos por la aplicación final. No vamos a decir cuál de estos métodos de evaluación es el más adecuado, ya que esto depende en gran medida del centro al que se oriente el *topic map*. En cualquier caso debemos indicar que deben realizarse estas evaluaciones con regularidad para asegurar la calidad del *topic map* a medida que este vaya creciendo; ya que, el *topic map* es una aplicación viva, en constante cambio a medida que se desarrolla con la colección y el conocimiento contenido en el dominio de la aplicación⁶⁹⁰.

5. Fase de documentación.

La documentación de un *topic map*, al igual que se hace con la documentación de cualquier programa informático, está orientada a distintos tipos de usuario⁶⁹¹. El diseñador, para el posterior mantenimiento del *topic map* cuenta con documentación diversa según a quién vaya orientada. Una documentación sencilla, orientada a la comunidad más general de usuarios cuenta con manuales en distintos idiomas y conservados independientemente del *topic map*; así como los documentos de *frequently asked questions* (FAQ) para cada perfil de usuario.

Sin embargo, una documentación mucho más completa acerca del *topic map* debe estar integrada en el mismo para que este pueda ser autodocumentado; es decir, que cualquier elemento de la estructura del mismo puede ser explicado desde dentro sin tener que acudir a documentos ajenos a él. Para ello, se asignan aquellos documentos explicativos a las *occurrences* de las clases e instancias que se desean documentar. Las clases *occurrence* nos permitirán modelar los distintos tipos de documentación, como puede ser si trata de una guía editorial, de una descripción o de una FAQ. Los contextos asignados a las *occurrences* permiten indicar la audiencia final a la que va destinada la documentación y su lengua, ya que es tarea del *topic map* mostrar la documentación correcta al usuario correcto y en su lengua.

⁶⁹⁰ GYUN OH, S. ; PARK, O. Design and user's evaluation of a topic maps-based Korean folk music retrieval system. En MAICHER, L. ; SIGEL, A. ; GARSHOL, L. *Leveraging the semantics of Topic Maps*. Berlin : Springer, 2007. pp.84-87.

⁶⁹¹ RATH, H. *The Topic Maps Handbook*. Germany : Empolis, 2003. p.35.

5.7. Herramientas de edición de *Topic maps*.

La tecnología de los *topic maps* se ha ido difundiendo progresivamente entre las empresas y algunas instituciones de manera progresiva en los últimos años, lo que ha supuesto que sus autores hayan creado herramientas específicas para tratar con este tipo de tecnología. Vamos a comentar algunas de las principales organizadas de acuerdo a sus siguientes funciones⁶⁹²:

- I) Diseño y edición de *topic maps*. El proceso de diseño depende de los recursos de información, por tanto es importante comprender el contexto y el significado real de la información con el fin de poder crear un *topic map* coherente. Así, que el documentalista, a partir de una colección documental identifica los *topics* y *topic types*, las *occurrences* y las *associations*. Si bien los editores de *topic maps* ya no obligan al conocimiento de la sintaxis XTM y sus especificaciones por parte del autor de *topic map*, debe tener un conocimiento general del mismo al igual que un diseñador de páginas web conoce algo de etiquetas XHTML. Sin embargo, dado que el etiquetado distrae al editor de cuestiones tales como: si el archivo XTM está bien formado, o acerca de la coherencia del documento. Es por esto que se utilizan los interfaces gráficos de usuario que hacen más intuitivo el proceso de edición, de forma que sea más fácil centrarse en la semántica de los documentos que en su sintaxis. De esta manera este tipo de software facilita la navegación por cualquier tipo de *topic map* además de facilitar la importación y exportación de *topic maps*, su modificación y actualización.
- II) Representación y navegación por *topic maps*. A este nivel, ya contamos con un *topic map* y lo que importa ahora es ser capaces de mostrar la información, así como el modo de acceso por parte del usuario a la colección representada por el *topic map* de la manera deseada por el documentalista. La manera en que se muestra la información debe ser clara para que el usuario comprenda cómo navegar por ella hasta satisfacer sus demandas informativas. Para el empleo de *topic maps* en navegación por

⁶⁹² HATZIGAIIDAS, A. ; PAPASTERGIOU, A. ; TRYFON, G. ; MARITSA, D. Topic map existing tools: a brief review. En *Proceedings of the international conference on theory and applications of mathematics and informatics- ICTAMI 2004*. En: <http://www.emis.de/journals/AUA/acta8/Hatzigaidas-Papastergiou-Tryfon-Maritsa.pdf>. Consultado el 15/12/2009.

grandes colecciones se requiere facilidad de uso y comprensión de interfaces de usuario. La interacción con el usuario, los distintos tipos de presentación de la información y de navegación, o la recuperación con búsquedas sencillas facilitan el proceso de localización de información si todo ello ayuda a la comprensión global y contextual de la dominio de conocimiento representado por el *topic map*. Esto supone que los usuarios deben ser capaces de centrarse en cualquier parte del *topic map* para la consulta precisa de información acerca de un *topic* concreto, proporcionando distintos niveles de detalle que les permita una navegación intuitiva.

- III) Visualización de *topic maps*. Para una representación exacta y por tanto una navegación eficiente por el *topic map* se hace necesario proporcionar una imagen conceptual de los documentos representados. La investigación en este campo tiene que ver con el desarrollo de estrategias orientadas a la visualización de recursos como medio de asistencia al usuario para la interpretación del conocimiento. Esto se consigue por medio de grafos y árboles jerárquicos, si bien lo más adecuado para los *topic maps* es el empleo de grafos consistentes de nodos y arcos. El problema surge cuando se tiene que representar un *topic map* con millones de *topics*. Munzner propuso el empleo de la geometría hiperbólica en vez de la euclídeana para la representación de un gran número de nodos en una misma ventana⁶⁹³. Los árboles hiperbólicos, las espirales o árboles cónicos son algunas de estas aproximaciones.

Otra solución a la falta de espacio en la pantalla es la representación de los *topic maps* en tres dimensiones. El empleo de la tercera dimensión permite al usuario un mejor control sobre la información a explorar, a modo de entorno en tres dimensiones que responde a los movimientos del usuario, acercando los *topic* más relevantes y alejando los menos relevantes, cambiando su punto de atención.

⁶⁹³MUNZNER, T. H3: *Laying out large directed graphs in 3D hyperbolic space*. En: <http://www.graphics.stanford.edu/papers/h3/>. Consultado el 15/12/2009.

Según Ahmed contamos con dos modos de representación de *topic maps*⁶⁹⁴, en perspectiva y en detalle. Cada una de estas técnicas es apropiada para un nivel de detalle diferente. Unas pueden representar eficientemente la totalidad del *topic map* y otras lo hacen con más detalle de cualquier parte del *topic map*.

- IV) Consulta de *topic maps*. Una de las características que dota a los *topic maps* de mayor capacidad para ofrecer a sus usuarios la información demandada es su capacidad para la búsqueda de otros *topic maps* o para la recuperación en bases de datos con sintaxis XTM. Análogamente a las consultas en bases de datos relacionales, es posible una recuperación estructurada de datos *topic map* por medio de lenguajes específicos de recuperación como es el lenguaje TOLOG o el lenguaje normalizado entre la comunidad de *topic maps*: TMQL. Estos lenguajes, como ya hemos visto, permiten la búsqueda de información pertinente no sólo con la resolución de las preguntas de sus usuarios sino que también es posible entender las consultas como un modo especial de navegación, especialmente si éstas se realizan sobre *associations*. Veremos qué herramientas hacen posible el empleo de las consultas con estos lenguajes sobre *topic maps*.

De acuerdo con las funciones hasta ahora descritas, vamos a pasar a describir los tipos de herramientas con las que contamos para la consecución de cada una de estas funciones. De acuerdo con Hatzigaidas vamos tres categorías de herramientas *topic map*⁶⁹⁵. La primera son los motores de *topic maps*. Un motor proporciona un interfaz de programación de aplicaciones comprensible que permite a los programadores crear y modificar estructuras *topic map*. Pueden incluso ser utilizados para la importación, exportación y consulta de *topic maps*. Además pueden ser empleados para la gestión, recuperación y mantenimiento de los *topic maps* almacenados en una base de datos. Aunque no se puede decir que la visualización sea parte de un motor, la mayoría de ellos proporcionan esta función como servicio de valor añadido para un mejor empleo de los *topic maps*.

⁶⁹⁴ AHMED, K. Topic maps for repositories. En *XML Europe Proceedings*, 2000. En: <http://www.gca.org/papers/xml europe2000/papers/s29-04.html>. Consultado el 15/12/2009.

⁶⁹⁵ Id.

La segunda categoría de herramientas es aquellas relacionada con la función de navegación. Los navegadores *topic maps* son herramientas que permiten al usuario ver y navegar a través de *topic maps*; para ello, emplean distintas técnicas que permiten al usuario alcanzar una visión comprensiva de los *topics* y sus interconexiones.

Finalmente, la tercera categoría son los editores de *topic maps*, herramientas que permiten al usuario crear y modificar *topic maps*, normalmente en entornos textuales.

Motores de *Topic maps*.

Ontopoly. Se puede descargar gratuitamente de la red⁶⁹⁶. Se necesita instalar un programa servidor apache-tomcat en el mismo ordenador desde el que se edita el *topic map*. Permite la navegación con un navegador genérico. Permite una rápida evaluación de cualquier *topic map* elaborado en XTM, o bien la elaboración sencilla de un *topic map* introduciendo los *topic*, *associations* y *occurrences*. Desde abril de 2009 el editor es open source, lo que permite mejorar las posibilidades del editor. Además, permite la interoperabilidad de los *topic maps* diseñados con este editor con otros editores. Razones por las cuales, como explicaremos en el capítulo acerca de la construcción del *topic map* para el CEPC, lo escogimos. Su funcionamiento es muy sencillo, por medio de formularios con los que se asignan los valores a las distintas propiedades y elementos del *topic map*. Un ejemplo de ello lo tenemos en la figura siguiente del *topic map* para el CEPC.

⁶⁹⁶ ONTOPOLY. En: <http://www.ontopia.net/solutions/ontopoly.html>. Consultado el 15/10/2009

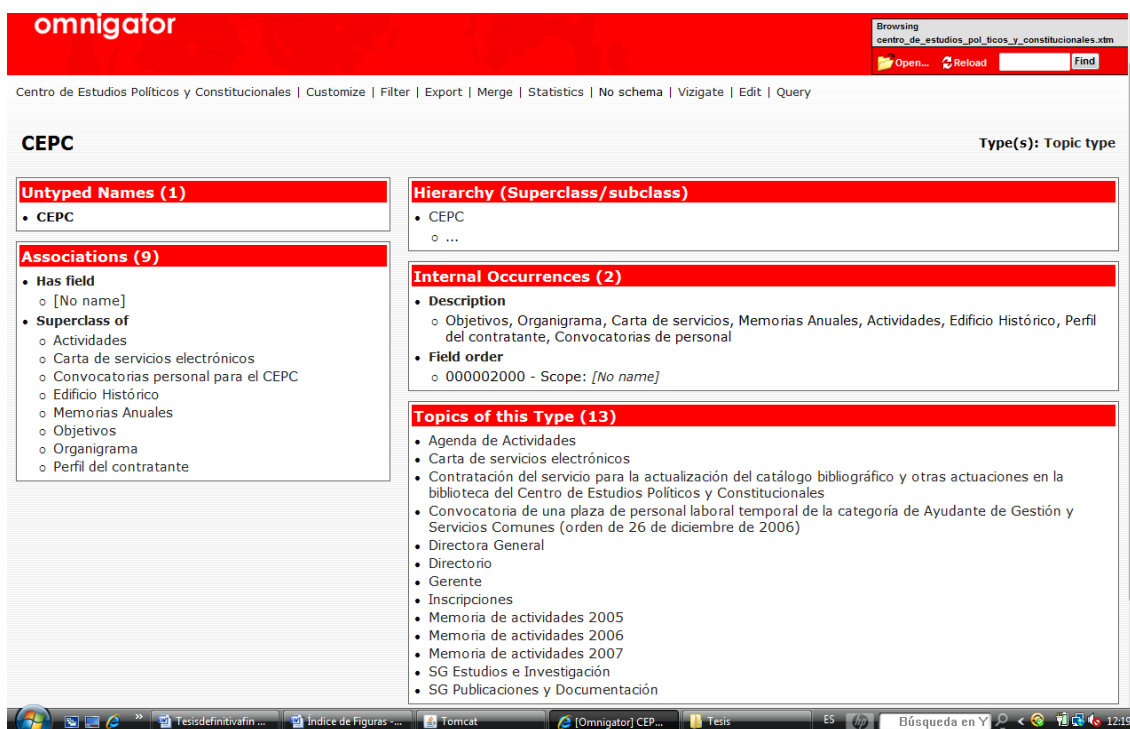


Figura 55. Muestra de diseño del *topic map* para el Centro de Estudios Políticos y Constitucionales.

TM4J. Se trata de un motor de *topic maps* de código abierto muy utilizado por la comunidad *topic map* tanto para la autoría, como para su edición y publicación. Fundamentalmente está constituido por los siguientes proyectos⁶⁹⁷: i) TM4JEngine. Motor de procesamiento *topic map* escrito en sintaxis Java, ii) TMNav. Es una aplicación de escritorio para la navegación en *topic maps*. Emplea un interfaz basado de usuario basado en grafos, iii) Pankoucke. Es una biblioteca para la representación abstracta con grafos a partir de *topic maps*, iv) TM4Web. Proporciona código de soporte y aplicaciones de referencia para la integración de su motor *topic map* con aplicaciones web.

tmproc. Motor *topic map* que ejecuta la norma ISO/IEC 13250 en sintaxis Python. Es interesante por haber sido el primer editor de *topic maps* disponible en red de forma gratuita. Sin embargo tiene el problema de no ser capaz de procesar los *topic maps* contruidos con el lenguaje normativo XTM 1.0, sino que sólo entiende python. La autoría de esta aplicación es de la empresa Ontopia⁶⁹⁸.

⁶⁹⁷ TM4J. En: <http://tm4j.org/about.html>. Consultado el 15/10/2009.

⁶⁹⁸ TMPROC. En: <http://www.ontopia.net/software/tmproc/>. Consultado el 15/10/2009.

SemanText. Aplicación desarrollada para demostrar cómo la norma ISO 13250:2000 puede ser empleada para representar redes semánticas⁶⁹⁹. Las redes semánticas son empleadas en aplicaciones del campo de la inteligencia artificial tales como motores de inferencia y sistemas expertos. La aplicación construye bases de conocimiento en forma de redes semánticas a partir de un *topic map*. Se puede incluir nueva información de forma automática definiendo reglas de inferencia; reglas que son definidas de acuerdo con la norma ISO 13250.

Tm4JScript. Se trata de un programa de código abierto⁷⁰⁰. El editor se basa en la ingeniería de un modelo orientado a objetos. Es un diseño muy sencillo que emplea índices y tablas de funciones (hashing) para operar, en el sentido de que con la función hash asigna una clave que es representación de un documento o conjunto de documentos. Con lo cual, más que trabajar con documentos se trabaja con claves. De esta manera, en general, se puede prescindir aquí de un lenguaje de consulta.

tinyTIM. Es un programa sencillo que opera sobre Java para la compilación de *topic maps*. Utiliza los interfaces de programación para el tratamiento de *topic maps* denominado TMAPI⁷⁰¹, que viene a ejercer el mismo papel que el modelo de objetos de documento (DOM) hace para XML⁷⁰².

TMAPI/PHP. Aplica una API para PHP sobre la base del proyecto TMAPI. Esta API permite a los desarrolladores de PHP la inserción sencilla y normalizada de *topic maps* en sus aplicaciones⁷⁰³.

xSiteable. Es un paquete de aplicaciones para el desarrollo de sitios creados en XSLT y disponible para varias plataformas y sistemas operativos. Tiene un sencillo lenguaje para la

⁶⁹⁹ SEMANTEXT. Descargable gratuitamente. En: <http://sourceforge.net/projects/semantext/>. Consultado el 15/12/2009.

⁷⁰⁰ Tm4Jscript. En: <http://tm4jscript.sourceforge.net>. Consultado el 15/12/2009

⁷⁰¹ Un interfaz TMAPI (Interfaz de programación de aplicaciones para *topic maps* es un conjunto de métodos que ofrece cierta biblioteca para ser utilizado por otro software como una capa de abstracción.

⁷⁰² Es decir, que es un modelo computacional a través del cual se permite a los programas acceder, añadir y modificar de forma dinámica el contenido estructurado de los *topic maps*.

⁷⁰³ TMAPI. En: www.tmapi.org. Consultado el 15/12/2009.

notación de contenidos denominado xSiteable Notation que emplea XTM para las estructuras y la gestión de los hipervínculos. Es gratuito y es descargable en la red⁷⁰⁴.

Navegadores de *Topic maps*.

The “V” *Topic map* Browser. El navegador “V” es un prototipo de navegador *topic map* en HTML. Está escrito en lenguaje Python y utiliza el kit de herramientas GooseWorks como servidor. Es gratuito⁷⁰⁵.

Panckoucke. Es una biblioteca Java de asistencia al desarrollo de aplicaciones *topic map* como herramientas de navegación. Se trata de un subproyecto de “tm4j”, empleado como motor *topic map* subyacente. Integra la aplicación “tmharvest”, otro subproyecto de tm4j, para la creación automatizada de *topic maps* a partir de datos estructurados. Panckoucke es libre y de código abierto con licencia Apache Software Foundation⁷⁰⁶.

TM4Web/Velocity. Es un paquete informático, basado en el proyecto TM4J, para el tratamiento de *topic maps* como páginas HTML. Permite su edición sin necesidad de estar conectado a internet, así como un control sobre el proceso de edición en el marco de una aplicación web que muestra el desarrollo de la muestra junto con plantillas de ejemplo para su publicación como página HTML⁷⁰⁷.

TM View. Basado en una aplicación denominada i-Disc desarrollada por Hofmann, Wendler y Froehlich. Es una herramienta para visualizar y explorar interactivamente *topic maps*. El sistema presenta una jerarquía de *topics* con un diseño plano radial, mediante la codificación de diferentes niveles de jerarquía, como anillos separados. Las *associations* se muestran por medio de la sintaxis de mapas conceptuales: arcs. Sin embargo, a causa de la separación de estas dos estructuras *topic map* en diferentes dimensiones espaciales, se hace difícil la consulta de *topic maps* complejos con este tipo de visualización gráfica⁷⁰⁸.

⁷⁰⁴ Xsiteable. En: <http://xsiteable.sourceforge.net/>. Consultado el 12/12/2009.

⁷⁰⁵ ALGERMISSEN, J. V *topic map* browser. En: <http://www.topicmapping.com/v.html>. Consultado el 12/10/2009.

⁷⁰⁶ TM4J. Subproyecto Panckoucke. En: <http://tm4j.org/x1k1pbb6tr-0.html>. Consultado el 15/12/2009.

⁷⁰⁷ TM4J. Subproyecto TM4Web/Velocity. En: <http://tm4j.org/tm4web-velocity-0.1.html>. Consultado el 15/12/2009.

⁷⁰⁸ ZANGERL, A. TM View. En: <http://search.cpan.org/~alphazulu/TM-View-0.5/>. Consultado el 12/01/2010.

HyperGraph. Se utiliza para la visualización de *topic maps*. Es un proyecto de código abierto en Java que permite trabajar con la geometría hiperbólica y, especialmente, con los árboles hiperbólicos. Lo hace proporcionando una API para visualizar la geometría hiperbólica, manejo de gráficos y distribución de árboles hiperbólicos⁷⁰⁹.

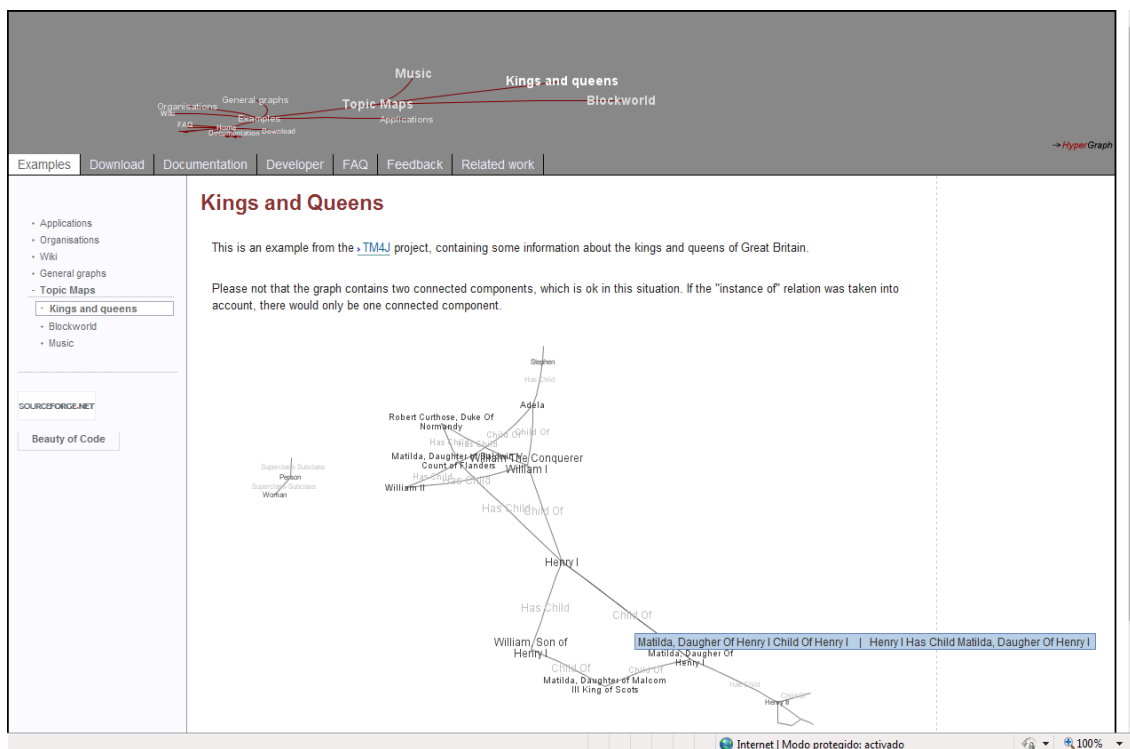


Figura 56. Ejemplo de *topic map* de reyes de Inglaterra creado con Hypergraph.

Thinkgraph. Es un programa de diseño 2D orientado a la generación de mapas conceptuales. Se puede decir que es una mezcla entre una aplicación de diseño 2D y un editor de mapas conceptuales. A diferencia de la mayoría de programas de diseño y de edición de mapas conceptuales, este programa no emplea un formato propietario sino que se apoya en dos normas XML: SVG para el modelado de la presentación del documento y XTM para el modelado de la parte de datos y relaciones del documento. Para visualizar los diagramas generados con este programa hace falta un visualizador de SVG⁷¹⁰.

⁷⁰⁹ HYPERGRAPH. En: <http://hypergraph.sourceforge.net/examples.html>. Consultado el 12/01/2010.

⁷¹⁰ THINKGRAPH. En: <http://www.thinkgraph.com/>. Consultado el 15/12/2009.

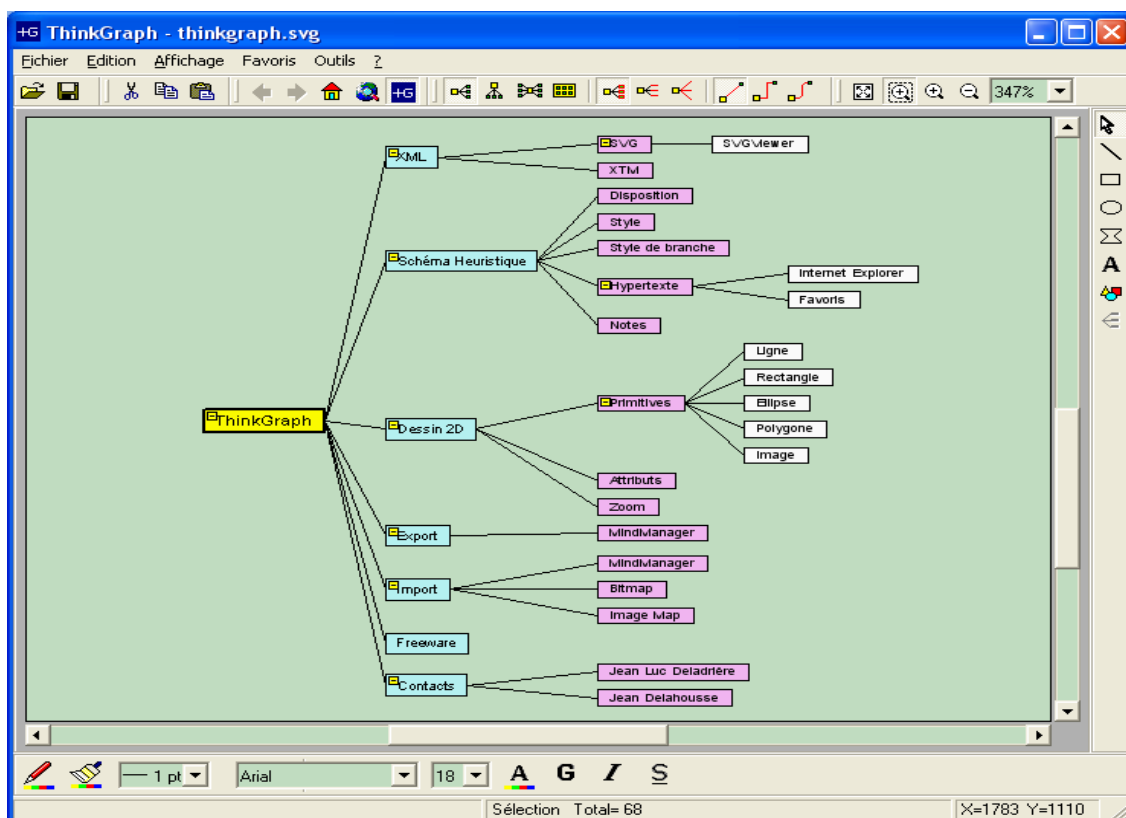


Figura 57. Diagrama generado con Thinkgraph.

Editores de *Topic maps*.

K-Discovery. Es un software para la gestión del conocimiento con *topic maps* en entornos cooperativos. Se trata de un marco gestor con *topic maps* constituido por tres módulos. Un motor *topic map* basado en groupware, una herramienta para su diseño y un navegador basados en groupware⁷¹¹.

ZTM. Tiene como objetivo permitir el desarrollo distribuido y mantenimiento de sitios web con herramientas de control semántico de contenidos como los *topic maps* por medio del tratamiento de la información con un modelo de datos derivado del modelo de datos para *topic maps* TMDM normativo de la ISO13250, para que el contenido de los sitios web pueda ser gestionado usando el lenguaje Zope. Esta aplicación integra navegadores y aplicaciones para la recuperación de *topic maps*⁷¹².

⁷¹¹ Se trata de una herramienta gráfica para desarrollar, configurar y mantener plantillas para *topic maps*.

⁷¹² ZTM *Topic maps*. En: <http://sourceforge.net/projects/ztm/>. Consultado el 12/12/2009.

Topic map Designer. Es un editor y un visor gráfico, todo en uno. Creado por Ronald Heckel, un profesor de Empresas de Informática de la Universidad Tecnológica de Dresde, como trabajo de tesis. Es gratuito y aplicable desde Windows. Se puede descargar en alemán e inglés, pero la actualización se está haciendo de la versión en lengua inglesa. Es un programa bastante didáctico acerca de los elementos que integran un *topic map*, e integran un tipo de visualización de árbol hiperbólico que se está aplicando como modelo de visualización en las actuales arquitecturas de información⁷¹³.

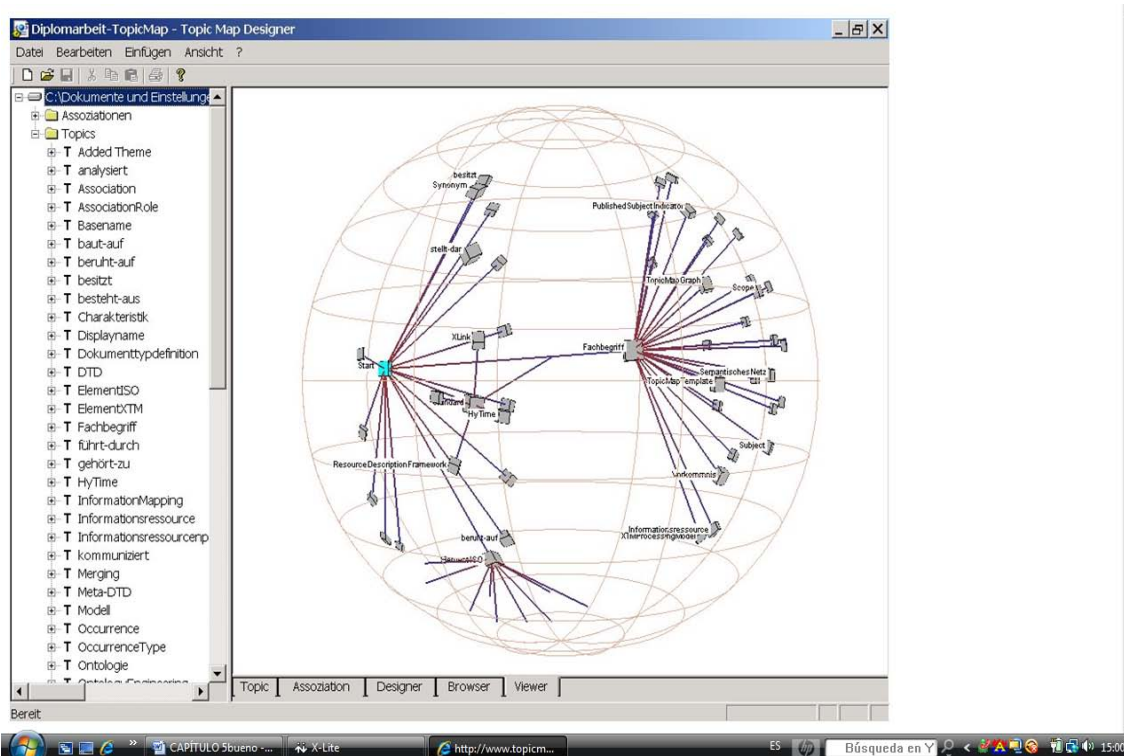


Figura 58. Ejemplo de *topic map* con la aplicación *Topic map designer*.

GNEWSYS. Knowledge Networking and Organization System es una aplicación web para el desarrollo y mantenimiento de contenidos para la web semántica. Desarrollado en Python, funciona como un producto instalado en Zope. Se trata de un software libre distribuido bajo licencia GNU GPL que permite su exportación a sintaxis de *topic maps* XTM⁷¹⁴.

Simple Topic maps Management. Se trata de un módulo en Perl para poder almacenar *topic maps* en una base de datos relacional. Es sólo una versión de prueba y es todavía un

⁷¹³ GEROIMENKO, V. ; CHEN, Ch. Op. cit.

⁷¹⁴ GNEWSYS: A kernel for semantic computing. En: <http://www.gnu.org/software/gnewsys/>. Consultado el 15/12/2009.

trabajo en desarrollo. Dispone de un interfaz web sencillo integrado para la edición de *topic maps*⁷¹⁵.

XSLT stylesheets. Se trata de hojas de estilo creadas por Nikita Ogievetsky de la empresa Cogitech para la transformación de DTD's del lenguaje XML a sintaxis XTM de *topic maps*. Asimismo, están diseñadas para la conversión de *topic maps* diseñados con las herramientas de código cerrado de empresas como Cogitech, Empolis, Infoloom, Ontopia y Techquila en sintaxis XTM. Estas hojas de conversión son gratuitas, al igual que lo es el programa de conversión creado por Mintert xtm2xhtml para convertir sintaxis XTM de un *topic map* en sintaxis XHTML interpretable por un programa navegador⁷¹⁶.

MDF. Se trata de un entorno en código abierto para tratar con metadatos ideado por Kal Ahmed, quien lo define como: una combinación de un método de edición de módulos reutilizables para el procesamiento de metadatos, y aplicaciones en lenguaje Java. El entorno ha sido diseñado para facilitar la incorporación de recursos, la extracción y filtrado de metadatos con los que generar automáticamente un *topic map*⁷¹⁷.

TMTab. Es un plugin para Protégé que permite el empleo de este editor de ontologías para la edición de *topic maps*. La funcionalidad actual del plugin se limita a la exportación de archivos XTM de una ontología desarrollada en Protégé, pero es lo suficientemente flexible tanto para permitir la asignación de propiedades a los conceptos definidos en la ontología como para la representación de *topic names*, *occurrences* o la definición de *associations* entre los conceptos o *topics*. La aplicación es gratuita y descargable desde la página principal de Protégé⁷¹⁸.

⁷¹⁵ LAMARCA LAPUENTE, M.J. *Hipertexto: El nuevo concepto de documento en la cultura de la imagen*. En: http://www.hipertexto.info/documentos/maps_tematic.htm. Consultado el 15/12/2009.

⁷¹⁶ OGIEVETSKY, N. *Building dynamic web sites with topic maps and XSLT*. En: <http://www.cogx.com/?si=urn:cogx:resource:tmwsites>. Consultado 15/01/2010.

⁷¹⁷ AHMED, K. (et al.). *Professional XML Metadata*. En: <http://www.techquila.com/mdf-dl.html>. Consultado el 15/01/2010.

⁷¹⁸ *TMTab*. En: <http://www.techquila.com/tmtab.html>. Consultado el 15/12/2009.

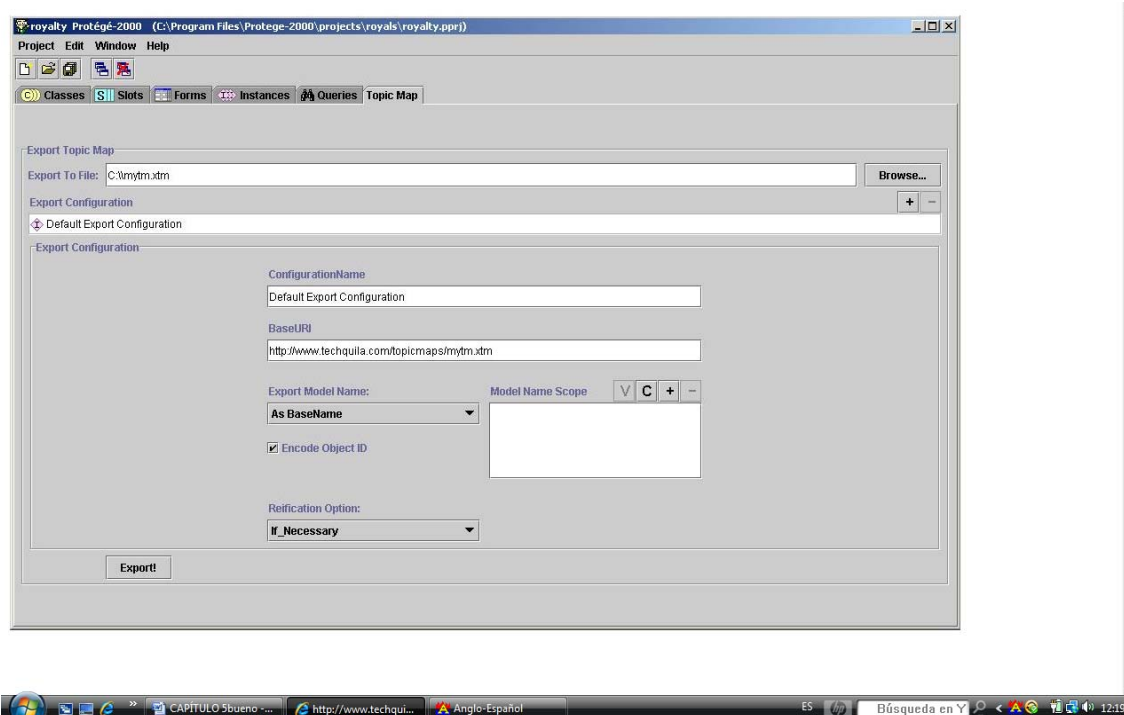


Figura 59. Ejemplo de formulario para la edición del basename de un *topic* con TMtab.

Proyecto Ceryle. Es una herramienta gratuita de asistencia al usuario para la organización de su información. Está diseñada para ayudarle a no perder la localización de los recursos informativos en repositorios de información personales. Incluye funciones para ayudarle a almacenar, localizar, y visualizar sus fuentes utilizando una visualización gráfica⁷¹⁹.

TM4L. Editor y Navegador. Es un entorno de edición para la creación, mantenimiento y uso de programas sobre el soporte de ontologías basadas en la norma ISO *topic maps*. El editor TM4L, basado en el software TM4J, es una herramienta de desarrollo *topic map* alternativo a los sistemas de edición convencionales destinados a la integración de los recursos de aprendizaje ya existentes en la web⁷²⁰.

Mapalizer. Es una herramienta de edición de *topic maps* escrita en Perl por Jan Algermeissen para la edición rápida de *topic maps*⁷²¹.

⁷¹⁹ ALTHEIM, M. Ceryle. En: <http://www.altheim.com/ceryle/>. Consultado el 15/12/2009.

⁷²⁰ DICHEVA, D. ; DICHEV, C. *Topic maps for e-learning*. En: <http://compsci.wssu.edu/iis/nsdl/index.html>. Consultado el 15/12/2009.

⁷²¹ ALGERMEISSEN, J. *Mapalizer*. En: www.topicmapping.com/mapalizer. Consultado el 15/12/2009.

Dtdoc. Se trata de una herramienta programada por Garshol para generar descripciones de tipos de documentos (DTD) para documentos en XTM. Es gratuita y opera sobre lenguaje Python⁷²².

5.8. Visualización con *Topic maps*.

En este epígrafe realizamos un análisis sobre la visualización en los interfaces de herramientas de edición *topic map* a fin de evidenciar la mejora que se produce en la navegación y visualización de los contenidos con el empleo de esta tecnología.

La tecnología *topic map* pretende aprovechar las aportaciones que la investigación en mapas conceptuales ha aportado desde los trabajos de Novak⁷²³. Sin embargo, debemos tener presente⁷²⁴:

1º) La organización de la información con tecnología *topic map* y su visualización en mapas conceptuales son dos caras de la misma moneda, ambos conceptos van íntimamente ligados pero son dos aspectos distintos. Un aspecto es cómo organizamos los contenidos y otro diferente es cómo la vamos a visualizar en un Interfaz Gráfico de Usuario (GUI).

2º) Se carece, hasta el momento, de un modelo normalizado de visualización de *topic maps*. No obstante, a partir del análisis descriptivo de la visualización en *topic maps* podemos hacernos una idea de los modelos de visualización a los que se tiende por su mejor representación de la idea del *topic map*.

Son muchos los trabajos acerca del concepto de *topic map* con sus elementos más representativos⁷²⁵, pero contamos con muy pocos trabajos acerca de su visualización, dada la novedad de la tecnología.

Hasta el momento, una introducción al campo de la visualización de los *topic maps*, sus objetivos y requisitos para su representación lo tenemos tanto en el manual clásico de Le

⁷²² GARSHOL, L. *Dtdoc*. En: <http://www.garshol.priv.no/download/software/dtdoc/>. Consultado el 15/12/2009.

⁷²³ NOVAK, J. D. ; CAÑAS, A. *The theory underlying concept maps and how to construct and use them*. En: <http://cmap.ihmc.us/Publications/ResearchPapers/TheoryCmaps/TheoryUnderlyingConceptMaps.htm>. Consultado el 15/12/2009.

⁷²⁴ GARCÍA MARTUL, D. Visualización en topic maps: tendencias y propuestas. En *JOTRI 2003*. Leganés : Departamento de Biblioteconomía y Documentación, 2003.

⁷²⁵ PEPPER, S. *The TAO of Topic maps: Finding the Way in the Age of Infoglut*, 2000. En: www.ontopia.net/topicmaps/materials/tao.html . Consultado el 16/12/2009.

Grand⁷²⁶ como en el trabajo de Geroimenko y Chen⁷²⁷. Según estos autores, los principales modelos en la visualización de *topic maps* son: a) Visualización estática mediante un índice de índices; b) Una visualización dinámica en árbol hiperbólico.

Cada uno de estos modelos de visualización ha sido aplicado a diversos casos prácticos en la edición de *topic maps*. Comentaremos un caso de cada uno de estos modelos: i) el navegador Ontopia Omnigator aplicado a la navegación entre una serie de recursos digitales sobre ópera por medio de un índice de índices⁷²⁸ y ii) el navegador *topic map* construido en Java por la empresa EMPOLIS para la navegación y recuperación de las actas presentadas en el Congreso Internacional de XML 2002⁷²⁹.

A partir de la exposición de las últimas tendencias, tanto en visualización de sitios web y especialmente de *topic maps*, como en las últimas tecnologías para la edición de *topic maps*, explicamos la interfaz gráfica de usuario (GUI) que planteamos para el *topic map* del CEPC con la herramienta Vizigator. Esta permite:

1º) visualizar siempre en el mismo espacio de pantalla toda la información global, de modo que el acceso a los sucesivos nodos no suponga dejar de visualizar la totalidad del *topic map*, sino que gire para que la necesidad informativa del usuario esté siempre focalizada y a su vez contextualizada.

2º) La combinación de la ventana marco en la que está inserto el *topic map* con otra en la que se visualizaría la información específica de cada nodo.

Las investigaciones sobre visualización en *topic maps* han partido de los trabajos acerca de los mapas conceptuales⁷³⁰ como esquema de representación. Los mapas conceptuales son empleados para estructurar la información con el objeto de identificar las interrelaciones entre conceptos dentro de un dominio⁷³¹. Su procedimiento es fijar los conceptos relativos a un dominio que, una vez organizados, pueden también ser relacionados de forma

⁷²⁶ LE GRAND, B. Topic map Visualization. En Park, J. *XML Topic maps*. Boston : Addison-Wesley, 2002. pp.267-281.

⁷²⁷ GEROIMENKO, V. ; CHAOMEI, Ch. *Visualizing the semantic web: XML-based internet and information visualization*. London : Springer, 2003.

⁷²⁸ PEPPER, S. *The Italian Opera Topic map*. En: <http://www.ontopia.net/operamap/index.jsp>. Consultado el 15/12/2009.

⁷²⁹ *XML Conference & Exposition 2002*. En: <http://62.231.133.220/xmlus02-stv/st.jsp?name=XML%202002>. Consultado el 15/12/2009.

⁷³⁰ ROVIRA FONTANALS, C. El editor de mapas conceptuales DIGIDOCMAP y la norma topic maps. *Hipertext.net*, nº3, 2005. En: <http://dialnet.unirioja.es/servlet/articulo?codigo=1176266>. Consultado el 15/12/2009.

⁷³¹ NOVAK, J. D. ; GOWIN, D. B. *Learning how to learn*. New York : Cambridge University Press, 1984. pp.16ss.

transversal con otros conceptos⁷³². Tienen asociada de por sí una sencilla representación visual, de forma que un mapa conceptual es una estructura de datos gráfica ordenada con nodos tipificados; de forma que, son realmente una herramienta metacognitiva que permite visualizar la articulación de los conceptos y expresar los elementos conocidos acerca de un concepto, para a partir de ahí, desarrollar estrategias de profundización.

Cada nodo tiene asignado un tipo de nodo, tiene un único identificador, y tiene un contenido. Un nodo puede estar asociado a otros nodos. Los enlaces son siempre direccionables en el sentido de unir dos nodos en cualquiera de sus sentidos, así las líneas flechadas o no flechadas son empleadas para representar visualmente los enlaces pero no debemos perder de vista que una de las principales características de los nodos es su propiedad de simetría (“David trabaja en la Universidad Carlos III” o “En la Universidad Carlos III trabaja David”).

En algunas aplicaciones, los enlaces se caracterizan por pertenecer a una clase. En la visualización, los mapas conceptuales son empleados para proporcionar un mapeo entre estructuras visuales y su infraestructura semiótica por medio de la metáfora icónica⁷³³. El mapeo puede ser efectivo para asegurarse que los atributos visuales de los nodos, y enlaces clasificados están en correspondencia uno a uno con las clases a las que pertenecen. Entre los atributos visuales que pueden ser utilizados para la representación de las propiedades de los elementos constitutivos de un *topic map* incluyen color, forma y espesor.

Los mapas conceptuales han sido aplicados fundamentalmente en el área de la educación, pero el diseño web y sobre todo la arquitectura de información han comenzado a utilizarlos como metáfora de representación de los resultados en la recuperación de motores de búsqueda. Y entre uno de estos usos se encuentra el mapa del web, servicio de valor añadido de los sitios web que facilita la navegación icónica entre los recursos del sitio. Un claro ejemplo de ello lo tenemos en el motor de búsqueda Kartoo. Este, no solo es capaz de mostrar en un mapa conceptual los resultados de una búsqueda, sino que también

⁷³² MOREIRO GONZÁLEZ, J.A. ; LLORÉNS MORILLO, J. ; MARZAL GARCÍA-QUISMONDO, M.A. Mapas Conceptuales, Topic maps y Tesauros. En *Jornadas de Tratamiento y Recuperación de la Información*. Valencia : Universitat de Valencia, 2002.

⁷³³ MARCOS, M.C. Elementos visuales en sistemas de búsqueda y recuperación de información. *Hipertext.net*, nº3, 2005. En: <http://www.hipertext.net/web/pag257.htm#13338>. Consultado el 15/12/2009.

agrupa los resultados con criterios de proximidad semántica como lo hacen los motores de recuperación por cluster⁷³⁴. Un ejemplo del mismo lo mostramos en la figura siguiente⁷³⁵.

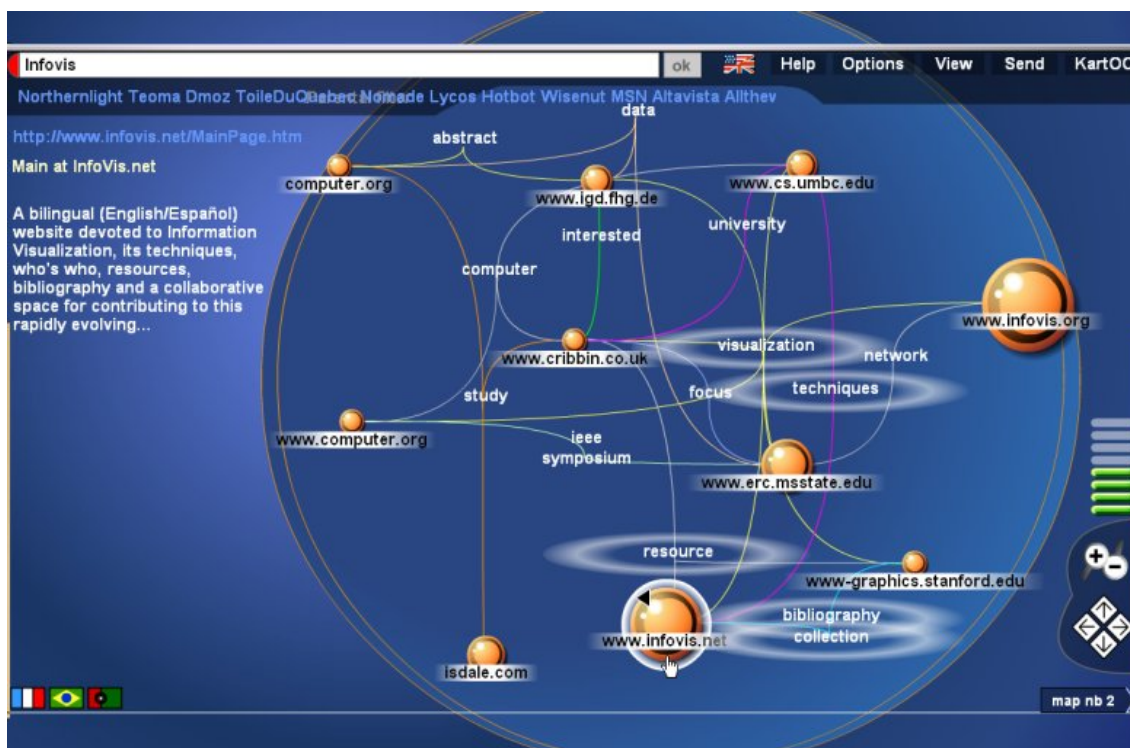


Figura 60. Presentación de los resultados de búsqueda con el mapa conceptual Kartoo.

Actualmente, los principales autores en visualización con mapas conceptuales provienen del Institute for Human and Machine Cognition⁷³⁶ fundado a partir de las ideas proporcionadas por uno de los principales impulsores de los mapas conceptuales: Novak. De hecho, algunas empresas tecnológicas que hacen un empleo intensivo de la organización del conocimiento, como la empresa fotográfica Kodak, han unido las posibilidades de la tecnología de los *topic maps* con la representación metafórica de los contenidos de su sitio web con mapas conceptuales para obtener un sitio navegable y extensible; es decir, que no obligue a la empresa a modificar la arquitectura de información a medida que se vaya incrementando el número de recursos representados⁷³⁷. Esto supuso un gran avance en su día para los *topic maps* dado que demostraba la capacidad de estos para

⁷³⁴ MARCOS, M. C. KartOO ahora más español. *El Profesional de la Información*, vol.12, nº2, 2003. pp.125-126. En: <http://www.elprofesionalde lainformacion.com/contenidos/2003/marzo/10.pdf>. Consultado el 15/12/2009.

⁷³⁵ KartOO. En: www.kartoo.com. Consultado el 15/12/2009.

⁷³⁶ IHMC. En: <http://www.ihmc.us/research/>. Consultado el 15/12/2009.

⁷³⁷ OGIEVETSKY, N. ; BADGER, T. Topic map solutions for Kodak digital camera accessories. En *XML Europe 2003*. En: http://www.idealliance.org/papers/dx_xml03/papers/02-03-02/02-03-02.pdf. Consultado el 15/12/2009.

permitir la navegación por miles de recursos informativos integrados en un repositorio multilingüe⁷³⁸.

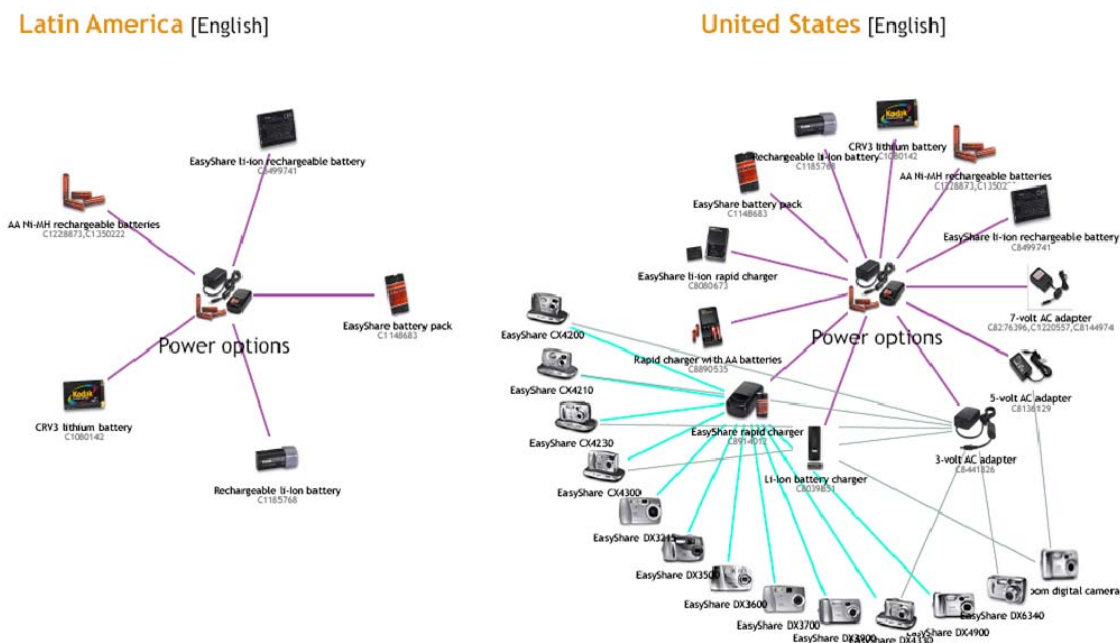


Figura 61. Presentación de los resultados de búsqueda con el *topic map* para Kodak.

5.8.1. Objetivos de la visualización.

El objetivo de la visualización de *topic maps* es ayudar a los usuarios a localizar información relevante rápidamente y facilitar la exploración de su estructura⁷³⁹. De acuerdo con esto, existen dos criterios a considerar en la visualización: representación y navegación⁷⁴⁰.

- i) La representación ayuda al usuario a identificar sitios pertinentes a sus demandas informativas. Su función es proporcionar al usuario, por medio de metáforas icónicas, una visión global del *topic map*. Visión global, donde se muestran las principales características de la arquitectura de información con el objeto de dotar a los usuarios de una visión global y contextualizada de los recursos de la biblioteca desde su perfil de usuario.

Para la representación, lo primero que debemos saber es el dominio del *topic* map, de qué trata. Una vez han sido identificadas las materias o clases de *topics*, necesitamos más

⁷³⁸ OGIEVETSKY, B. Tabular Topic maps. En *XML Europe 2003*. En: <http://web-services.gov/TTMSBIR20031016.ppt>. Consultado el 15/12/2009.

⁷³⁹ LE GRAND, B. Topic map visualization. En Park, J. *XML Topic maps*. Boston : Addison-Wesley, 2002. p.270.

740 *Id.*

información estructural, tal como la generalidad o especificidad del *topic map*. Este tipo de información se representa para ayudar a los usuarios a comparar distintos *topic maps* rápidamente y explorar sólo los más relevantes. La posición de los *topics* en el interfaz refleja su proximidad semántica. Estas propiedades pueden ser deducidas de la computación de métricas *topic maps*⁷⁴¹.

Por otra parte, los *topic maps* son bases de conocimiento multidimensional. Los *topics*, *associations* y *ocurrences* están caracterizados por muchos parámetros que aparecen de algún modo en la visualización, pero debemos tener presente que no es posible mostrar simultáneamente información general y en detalle de la misma. Podemos compararlo con un mapa geográfico. Así por ejemplo, un mapa del mundo no puede ser preciso más que hasta cierta escala. Si un usuario está interesado en los detalles, debe precisar el área de interés, por ejemplo escogiendo un país específico. Como en los mapas geográficos, necesitamos proporcionar distintas escalas de representación en *topic maps*. Esto obliga a que las visualizaciones sean dinámicas, para poder adaptarse a las necesidades de los usuarios en tiempo real. La combinación de tiempo y espacio ayuda al usuario a comprender la metáfora representativa del *topic map* junto con la ayuda de su experiencia sobre el mundo real.

ii) La navegación permite a los usuarios explorar el *topic map* y acceder a la información contextualizada⁷⁴². La navegación debe ser intuitiva con el objeto de que el usuario pueda ir de un punto a otro de la Web haciendo uso de sus capacidades cognitivas y de la memoria para la asociación de conceptos, de forma que es el propio usuario quien a través de las asociaciones desambigua la información recuperada y “reconstruye” su propio contexto, o semántica de mundo.

Para explicar esto vamos a recurrir a la metáfora de un viajero que lleva un GPS y un mapa. El viajero puede navegar en avión con lo cual tendrá una visión global y muy general del paisaje. Sin embargo, si viaja caminando por un bosque tendrá una visión muy específica y saturada de elementos del paisaje. Por tanto, la navegación es esencial ya que ayuda a los propios usuarios a construir su propio mapa cognitivo, entendiendo el mapa como la

⁷⁴¹ LE GRAND, B. ; SOTO, M. Topic maps metrics. En *Proceedings of Knowledge Technologies 2001*. En: <http://www2.gca.org/knowledgetechnologies/2001/proceedings/LeGrand%20Slides.pdf>. Consultado el 15/12/2009.

⁷⁴² LE GRAND, B. Op. cit.

representación cognitiva de un entorno, de forma que según navegues así percibirás el paisaje y así será la lectura e interpretación del mapa. Pero independientemente de que la navegación condicione la lectura e interpretación de los mapas cognitivos, lo que si permite es incrementar la capacidad de asimilar y comprender la información.

La navegación libre debe ser mantenida para pequeñas estructuras o usuarios expertos ya que la probabilidad de perderse es muy elevada. Para los que se inician en la navegación, aquellos que desconocen dónde comenzar la exploración, los caminos de navegación ya trazados son necesarios hasta que el usuario sea capaz de identificar los *topics* de su interés⁷⁴³.

Por lo tanto, necesitamos representar la totalidad del *topic map* con el objeto de ayudar a los usuarios a comprenderlo globalmente. Esta visión global debe reflejar las principales propiedades de la estructura. Sin embargo, los usuarios deben ser capaces de centrarse sobre cualquier parte del *topic map* y ver todas las dimensiones que necesitan. Siempre que muchas escalas requieran el uso de distintos niveles de detalle. Finalmente, los usuarios deben ser capaces de navegar fácilmente e intuitivamente entre estos distintos niveles de detalle. Por lo tanto la cuestión es saber qué técnicas de visualización satisfacen nuestras necesidades y pueden ser empleadas para representar *topic maps*.

El valor de cualquier sistema de información está condicionado por la calidad y cantidad de información contenida, pero al mismo tiempo por su facilidad o *findability* para encontrar dicha información⁷⁴⁴.

El estudio de la problemática del acceso a la información distingue dos modelos en función del sentido en el que se produce dicho acceso: *Push* y *Pull*.

En el modelo *Push* o de filtrado de la información, el sistema regularmente informa al usuario sobre aquellos documentos de nueva incorporación cuya temática se corresponde

⁷⁴³ BENGOCHEA, L. ; PATRICIO, M. A. Sistemas de visualización para bibliotecas digitales. *Revista Española de Documentación Científica*, vol. 28, nº3, 2005. p.275.

⁷⁴⁴ HASSAN MONTERO, Y. Visualización y Recuperación de información. En *II Encontro de Ciencias e Tecnologías da Documentacao e Informacao*, Vila do Conde, 2006. En: http://www.nosolousabilidade.com/hassan/visualizacao_y_recuperacion_de_informacion.pdf. Consultado el 15/02/2010.

con el perfil del usuario. Vendría a ser el modelo de la alerta informativa, una estrategia pasiva y no interactiva de acceso a la información por parte del usuario.

En el modelo *Pull* o de recuperación de información, es el usuario quien de forma activa resuelve sus necesidades de información. Para ello puede emplear dos estrategias: *querying* y *browsing*.

En el *querying* el usuario introduce en el sistema una estrategia de búsqueda compuesta por palabras clave que describen la demanda informativa del mismo. A continuación el sistema le devuelve una lista de resultados más o menos pertinentes con la demanda formulada. Este sistema permite la recuperación exhaustiva de documentos pero resulta poco preciso cuando el usuario desconoce las materias de la información buscada.

En la estrategia de recuperación por *browsing*, el usuario explora visual y espacialmente el conjunto documental, sin necesidad de tener que expresar por medio de un lenguaje de consulta cuál es su demanda de información.

En función del grado de especificidad de las necesidades de información del usuario podemos diferenciar entre varios tipos de *browsing*⁷⁴⁵: a) de búsqueda directa, donde el usuario sabe lo que está buscando; b) de propósito general, donde el usuario explora fuentes con una elevada probabilidad de resultar de interés; c) fortuito o por azar.

El ejemplo más claro de *browsing* es la navegación hipertextual, en la que el usuario explora visual y espacialmente el conjunto documental: reconoce visualmente la información buscada o aquellos enlaces que cree le llevarán a esa información, y se desplaza por este espacio de información saltando de un enlace a otro.

Entre los sistemas de información que ofrecen la posibilidad de búsqueda por *browsing*, cabe destacar aquellos que proveen de un medio específico para realizar *browsing* gráfico a través de las interfaces visuales de recuperación de información, VIRI. El *browsing* gráfico es similar al hipertextual, pero mientras que el hipertextual únicamente permite explorar los

⁷⁴⁵ COVE, J. F. ; WALSH, B. C. Online text retrieval via browsing. *Information Processing and Management*, vol. 24, nº1, 1988. pp.33-34.

documentos por separado, el *browsing* gráfico posibilita además visualizar conjuntos de documentos de forma global a través de su abstracción gráfica y no necesariamente textual.

Los VIRIs no sólo tienen el objetivo de ayudar al usuario a navegar y recuperar información, sino también el de ofrecer una visión global del conjunto documental. A través de esta vista global el usuario podrá predecir qué puede encontrar, e incluso adquirir nuevo conocimiento mediante la interiorización de la visualización⁷⁴⁶.

Desde una perspectiva teórica se pueden clasificar los VIRI de acuerdo con los siguientes factores: tipo de datos y de relaciones entre los datos, tipo de tarea, tipo de interactividad, destreza del usuario y contexto de uso⁷⁴⁷. Sin embargo con el objetivo de posibilitar el *browsing* gráfico a partir de un *topic map*, un VIRI debe cumplir las siguientes funciones básicas:

- Resumir la información con el objetivo de poder ofrecer vistas globales del conjunto documental.
- Resaltar la información más relevante a fin de facilitar al usuario la visualización del interfaz y así permitirle discernir la relevancia potencial de la información que consulta.
- Relacionar la información semejante o similar con el objetivo de permitir al usuario comprender y adquirir nuevo conocimiento a partir de las relaciones representadas entre los elementos visuales del VIRI.
- Recuperar la información a fin de que el usuario pueda finalmente satisfacer sus necesidades de información.

⁷⁴⁶ Id.

⁷⁴⁷ PFITZNER, D. ; HOBBS, V. ; POWERS, D. A Unified Taxonomic Framework for Information Visualization. En PATTISON, T. ; THOMAS, B. (eds.). *Proceedings Australian Symposium on Information Visualisation*. Adelaide : Conferences in Research and Practice in Information Technology, 2001. En: <http://crpit.com/confpapers/CRPITV24Pfitzner.pdf>. Consultado el 15/02/2010.

5.8.2. Modelos generales en visualización de información para topic maps.

Hoy en día existen gran variedad de técnicas de visualización para la representación de datos complejos. Entre ellos, los grafos, árboles y mapas, que como veremos son estos últimos las metáforas más adecuadas para la visualización de *topic maps*.

5.8.2.1. Grafos y árboles.

La representación de los *topic maps* como grafos parece natural mientras puedan ser vistos como una red de *topics* y *associations*. Aquí, el *topic map*, es mostrado como una gráfica compuesta de nodos y arcos. Sin embargo, esta simple representación puede volverse desordenada y difícil de interpretar cuando el tamaño de la estructura se incrementa. Como puede contener millones de *topics* y relaciones de todo tipo entre ellos, se vuelve necesario emplear técnicas de visualización gráfica muy sofisticadas.

Los grafos y árboles son adecuados para la representación de la estructura global de un *topic map* por ser capaces de reflejar las características de navegación guiada y contextual de los mismos. Sin embargo, los árboles son mejor percibidos por los usuarios cuando están jerarquizados y además son más fáciles de interpretar que las gráficas. En cambio, los *topic maps* no son jerarquías y por lo tanto no pueden ser representados directamente como árboles. Esto no impide que pueda ser interesante transformar pequeñas partes de los *topic maps* en árboles. Haciendo esto sobre una pequeña parte del *topic map*, para impedir la dispersión, podemos beneficiarnos de las ventajas de los árboles.

El desafío de la visualización con grafos es proporcionar gráficas que muestren nodos interrelacionados y contextualizados. Una primera solución es el empleo de la geometría hiperbólica, en vez de la geometría euclídeana, por la posibilidad de mostrar un gran número de nodos en pantalla.

Por otra parte, la calidad de la visualización puede verse mejorada por una disposición eficiente de los nodos, lo cual hace posible extraer información a partir de la distancia semántica entre nodos. Por ejemplo: los *topics* enlazados por una relación de asociación pueden ser representados contiguos entre sí, o bien los *topics* del mismo tipo o indicación a las mismas ocurrencias pueden ser agrupadas en clusters semánticos.

La visualización de *topic maps* mediante grafos y árboles tiene la cualidad de cumplir con una de nuestras primeras exigencias, que sean capaces de representar la totalidad del *topic map*. Sin embargo, los usuarios también necesitan ser capaces de ver información detallada acerca de los *topics* sobre los que se tiene interés. Este segundo requisito, que consiste en la representación de los distintos parámetros de un *topic map*, es la mayor dificultad que nos encontramos al representar una red semántica sobre materias. Distintas formas y colores son empleados para simbolizar las distintas dimensiones de nodos y enlaces de la gráfica. Este tipo de gráfica podría ser usada para visualizar un *topic map*; los *topics* serían nodos y las asociaciones podrían ser los enlaces. Sin embargo, el número diferente de formas, colores, iconos y texturas es limitado, y esto hace que este tipo de representación no sea la más conveniente para un *topic map* que represente millones de *topics* y asociaciones.

Con el objeto de mostrar la información detalladamente, es necesario centrarse en parte de un *topic map*. El grafo, como visión global del *topic map*, ayuda a los usuarios a comprender la estructura globalmente y a seleccionar *topics* específicos que puedan ser de su interés. Una vez que un *topic* ha sido seleccionado, es fácil mostrar información muy precisa acerca de la misma, pero se pierde el contexto que es el problema que en teoría resuelve la tecnología de *topic maps*.

Otro modo útil de representar un *topic map* es mostrar una lista o índice desde el cual sea posible seleccionar un *topic* y ver la información relacionada. La navegación es normalmente la misma que en los sitios web: son los usuarios quienes pinchan sobre el enlace con el que quieren abrir un nuevo *topic* o *association*. Sin embargo, hasta el momento, no ha sido posible una visualización global y precisa al mismo tiempo. Es fundamental que sean los propios usuarios quienes determinen durante la navegación el nivel de detalle de la información buscada. Para ello, hoy en día contamos con muchas herramientas que proporcionan visualizaciones gráficas interactivas.

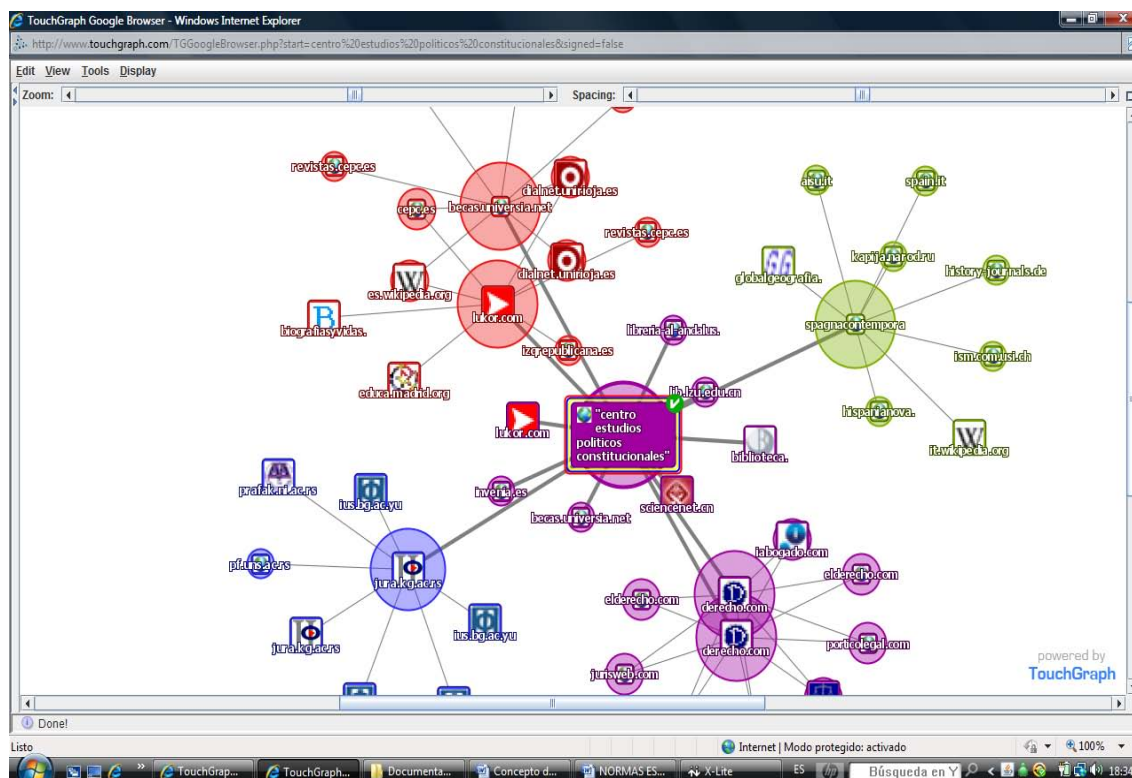


Figura 62. Representación visual de la información para el sitio web del CEPC con el editor de mapas conceptuales Touchgraph.

5.8.2.2. Mapas.

Los *topic maps* pretenden una mejora de la navegación dentro de colecciones de datos complejas. En el mundo real, los mapas son empleados para alcanzar esta meta. Es por tanto natural considerar la representación de los *topic maps* como mapas.

Un mapa representaría los elementos más significativos de la estructura. La localización y el tamaño de estos elementos en el mapa resultan esenciales. En los mapas arbóreos, los datos son reorganizados con el objeto de reflejar la importancia relativa de los *topics*. Estos mapas arbóreos pueden ser empleados para representar los *topic maps*.

Uno de los desafíos de la visualización de los *topic maps* como mapas es encontrar las coordenadas óptimas para los *topics*. El algoritmo de mapas autogestionados puede ser empleado para alcanzar ésto por organización de los *topics* en una malla bidimensional en la que los *topics* relacionados aparezcan próximos entre sí.

El factor de análisis puede también ser empleado para computar las coordenadas *topic*. Un ejemplo de esto es el algoritmo de escala multidimensional (MDS) emplea medidas de similitud entre *topics* para proporcionar un mapa en dos dimensiones de la estructura.

Otros programas, proporcionan distintos tipos de mapas parecidos a mapas topográficos, con montañas y valles. En ellos es esencial el concepto de capa. Aquellos documentos que con contenido similar se sitúan en lugares próximos. En las cimas se muestran las concentraciones de documentos estrechamente relacionados; el número más elevado de documentos crea picos más altos. Los valles entre picos pueden ser interesantes porque contienen unos pocos documentos y de más contenido singular. Las etiquetas reflejan dos o tres *topics* principales representados en un área dada del mapa, proporcionando una rápida indicación acerca de lo que tratan los documentos. Las etiquetas adicionales a menudo aparecen cuando acercamos el mapa para tener un mayor detalle. Podemos aumentar el mapa a distintos niveles de tamaño para amplificar el mapa y revelar documentos y etiquetas adicionales. Este tipo de visualización permite la combinación de distintas representaciones en muchas ventanas, así los usuarios pueden escoger de acuerdo a su perfil de demandas informativas.

Las herramientas visuales para la localización de información permiten mostrar la relación entre documentos de acuerdo con unas propiedades a destacar, permitiendo a los usuarios explorar interactivamente la colección y descubrir rápidamente nuevos modelos de significado, tendencias y relaciones. Estas, pueden utilizar paisajes animados 3D que aprovechan la capacidad del hombre para la navegación en espacios tridimensionales, identificación de modelos conceptuales, o la comparación de documentos de características diferentes con criterios temáticos. Así, los usuarios pueden tener un control completo sobre la apariencia de la información. Las técnicas de realidad virtual incluyen interacción y empleo de distintos niveles de detalle. Asimismo, la inmersión en el mundo virtual hace que los usuarios se sientan más implicados en la visualización. Un ejemplo de ello podría ser la representación de los *topic maps* como ciudades virtuales donde los *topics* se representan como edificios cuyas coordenadas son computadas a partir de una matriz de analogías entre *topics*. Los usuarios pueden navegar libremente o seguir un tour guiado a través de la ciudad. Las propiedades de los *topics* son simbolizadas por las características de los correspondientes edificios, tales como un nombre, color, altura, profundidad, etc. Las *occurrences* y *topics* asociados son mostrados en dos ventanas en la parte inferior de la

pantalla. Como los usuarios están acostumbrados a la bidimensionalidad, un mapa bidimensional también se muestra en la misma pantalla. Ambas visualizaciones, el mapa y la ciudad virtual, son siempre coherentes. Los usuarios, pueden explorar el mundo e interactuar con sus datos. Sin embargo, pueden llegar a perderse en el mundo virtual. Para evitar este tipo de problemas se proponen una serie de vías para la navegación predefinida. Los distintos niveles de detalle hacen posible mostrar muchas escalas; los detalles aparecen sólo cuando el usuario está próximo a la materia de su interés.

Los *topic maps* cuentan con muchas capacidades pero pueden llegar a ser muy complejos visualmente. La visualización intuitiva de los interfaces de usuario puede reducir significativamente la carga cognitiva de los usuarios cuando trabajan con estas estructuras complejas. La visualización es una técnica prometedora para la mejora por parte de los usuarios de la percepción de la estructura en grandes espacios de información y para proporcionar facilidades en la navegación. Además, permite al usuario emplear una herramienta natural de observación y procesamiento, tanto a través de sus ojos como de su cerebro, para extraer conocimiento más eficientemente.

5.8.3. Modelos de visualización propios de *Topic maps*.

Antes de explicar los principales casos de visualización encontrados en las aplicaciones más conocidas de *topic maps*, vamos a explicar brevemente los fundamentos de su visualización.

En visualización para *topic maps* se debe partir de dos ideas claras⁷⁴⁸:

1. Cada *topic map* presenta dos niveles: de *topics* y de *occurrences*.
2. El *topic map* puede asumir tres tipos de estructuras:
 - a. Lineal: Una línea recta, o un círculo en caso de tratarse de una estructura lineal cerrada, donde los *topics* están interrelacionados a través de *associations* secuenciales; es decir, para la comprensión de un *topic* debemos pasar obligatoriamente por otro anterior.
 - b. En árbol: Creando una jerarquía que permite pasar de un nodo superior a otro dependiente.

⁷⁴⁸ OLIVEIRA, A. Topic Maps na visualização de informação no ensino e treino. En *Proceedings da Associação Portuguesa de Sistemas de Informação*. Guimarães : Conferência da Associação Portuguesa de Sistemas de Informação, 2001. En: <http://en.scientificcommons.org/20329751>. pp.13-14. Consultado el 15/02/2010.

- c. Red multirecursiva: Permitiendo la creación de una red de *topics* interrelacionados que permita una navegación recursiva entre los mismos. Por ejemplo una esfera sobre la cual se sitúen los *topics* sobre su superficie y las *occurrences* en su interior. Este modelo, del que más adelante veremos un ejemplo, nos permite dotar a la visualización del *topic map* de dinámica, girando la esfera hasta situar los *topics* demandados por el usuario de manera contextualizada tanto en relación con otros *topics* como en relación con sus *occurrences*⁷⁴⁹.

5.8.3.1. Modelo de visualización en árbol hiperbólico.

Esta versión de *topic map* contiene las actas del Congreso de XML 2002, creado por la empresa Empolis. Consiste en páginas HTML que representan el *topic map* y las comunicaciones en él presentadas, las cuales pueden ser visualizadas con un navegador web corriente.

Esta aplicación *topic map* gestiona y organiza en Web el acceso a las actas del Congreso, pero en un sentido amplio ya que podemos no sólo consultar cada una de las comunicaciones sino que además permite la consulta de *topics* tales como el lugar de lectura de la comunicación, el resumen, la fecha o el lugar de exposición de la misma, etc. Como tales *topics* presentan relaciones con otros *topics* y cada una de estas relaciones está tipificada, las relaciones que se comprenden son: “lleva a”, “está localizado en”, “cubre”, “tiene palabra clave”, “es presentado por”, “es mencionado en”. Cuando se llega al acta concreta de un documento particular, tal como es el caso de la comunicación de Pepper titulada⁷⁵⁰ “The TAO of *Topic maps*”, cuyo *topic* raíz es la comunicación presentada en este congreso, siendo designado el nombre del *topic* por una cadena de caracteres que es el título de la comunicación. A partir de este *topic* raíz, como las ramas de un árbol, se despliega un abanico de once *topics*, de los cuales cinco de ellos no tienen definida el tipo de relación de la que se trata por apuntar a *topics* que son los cinco elementos informativos que se proporcionan de cada una de las comunicaciones: acceso directo a la comunicación “paper”, nivel de presentación, hora y día de la presentación y un resumen de la comunicación. Los otros seis enlaces tienen designado su tipo y apuntan a *topics*

⁷⁴⁹ LU, Y. ; HORNUNG, J. *3D visualization of knowledge domains in learning environments*. En: www.editlib.org/index.cfm/files/paper_11876.pdf. Consultado el 15/02/2010.

⁷⁵⁰ PEPPER, S. The TAO of Topic Maps. Finding the way in the age of infoglut. En *XML Europe 2000*. En: <http://www.gca.org/papers/xml europe2000/pdf/s11-01.pdf>. Consultado el 10/02/2010

relacionados con el artículo a través de la relación palabra-clave; es decir, las comunicaciones están relacionadas entre sí fundamentalmente a través de las palabras clave y de los autores, siendo a su vez cada nombre de autor un *topic* del que parten otros *topics* como lugar de trabajo, organización para la que trabaja, datos biográficos y datos de contacto.

Sus elementos constitutivos son los siguientes:

I. Home. Página de inicio del *topic map*. Se encuentran enlaces al *topic* de inicio XML2002 y al Meta Index que está por debajo. Permite navegar por la versión más actualizada, explorar gráficamente el *topic map* en el árbol hiperbólico, denominado Star Tree, o el comienzo de una nueva búsqueda.

II. Meta Index. Es el lugar de entrada al *topic map*. Aquí se expone una lista con todas las clases de *topics*, se trata del Índice de índices. El índice de relaciones ilustra los tipos de relaciones entre los *topics*. Después siguiendo uno de los enlaces del índice, se logra acceder a todos los *topics* y relaciones modeladas en el *topic map*. Estando en el nivel *topic*, se puede ver la clase de *topic*, sus fuentes y *topics* relacionados.

III. Star Tree. Enlace, online, a la visualización gráfica del *topic map* comenzando por el *topic* "XML 2002".

Las opciones de búsqueda con las que se cuenta en este *topic map* son:

I) Una de ellas es la búsqueda online. Esta proporciona acceso contextual a todas las actas, finalidad del *topic map*, y a algunos sitios web seleccionados proporcionando información marco acerca de XML. Los motores de búsqueda automáticamente ordenan los resultados por medio de palabras clave, títulos y resúmenes de los resultados a texto completo. Esto permite que no sea necesario restringir la búsqueda a una cierta zona de texto. La estrategia de búsqueda es un caso independiente y el análisis de los términos reduce todas sus flexiones gramaticales a su forma base, por lo tanto no tenemos que preocuparnos si el término está en singular o plural, etc. Por otra parte, un sistema gestor controla la ordenación de los resultados de la consulta, así como presenta los sinónimos de los

términos y un esquema de clasificación, el cual relaciona por ejemplo “metadatos” con “vocabulario” o con “Dublin Core”.

II) Empleo del árbol en estrella⁷⁵¹. El árbol en estrella es una visualización gráfica e interactiva del *topic map*. El grafo *topic map* es representado como un árbol que es proyectado sobre una bola plana en cuanto que desaparece el borde exterior. Se puede mover el árbol hiperbólico sin más que pinchar en cualquier lugar, lo cual lleva desde el punto seleccionado en el centro o bien arrastrando el ratón. Se comienza con un nodo raíz y se puede expandir interactivamente hasta alcanzar todos los nodos en la exploración del *topic map*.

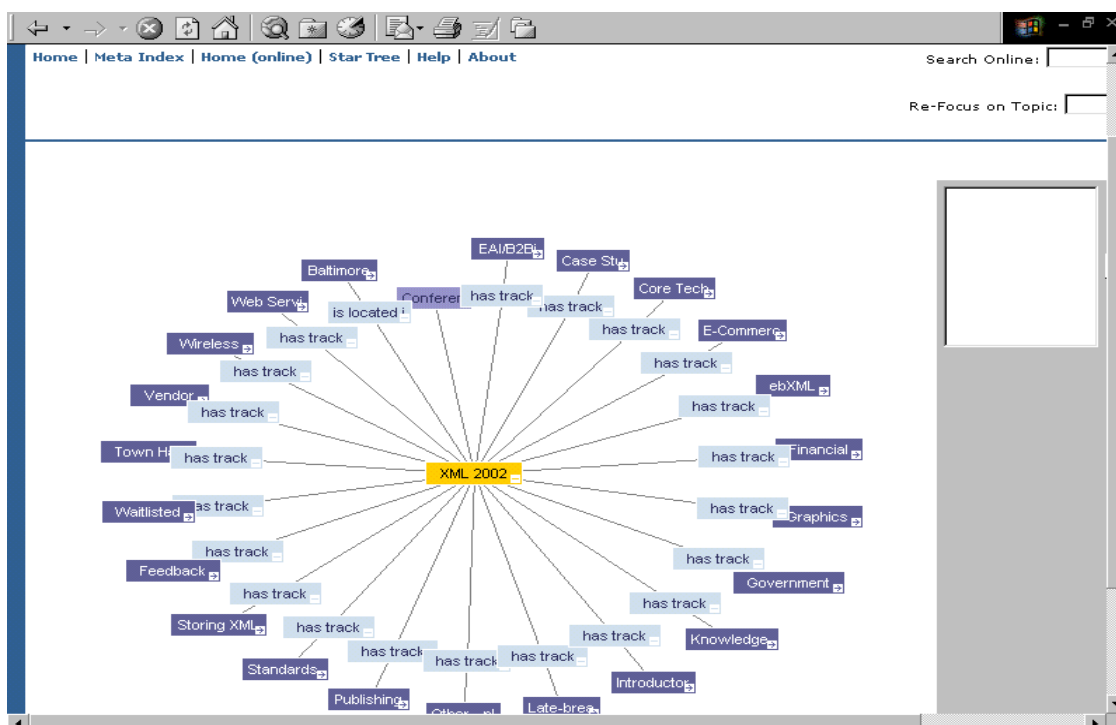


Figura 63. Representación en árbol.

La interacción del usuario con esta interfaz gráfica de árbol hiperbólico se efectúa por medio de los siguientes elementos:

XML 2002. Se trata del epígrafe raíz de todo el árbol entorno al cual giran el resto de *topics* relacionados entre sí mediante unas relaciones concretas previamente definidas.

is located in. Indica que existe una relación entre dos *topics* y además indica el tipo de relación de la que se trata. La lectura se lleva a cabo desde el nodo raíz hacia el nodo

⁷⁵¹ XML Conference 2002. En: <http://62.231.133.220/xmlus02-stv/st.jsp?name=XML%202002>. Consultado el 02/12/2009.

exterior. El árbol estrella cambia automáticamente la etiqueta cuando se visualiza la relación en orden inverso desde el nodo exterior hacia el nodo raíz.

Core Technologies. Se trata de una etiqueta *topic*. Para visualizar sus relaciones y ocurrencias se pincha en la flecha del margen inferior derecho. El tamaño del nombre *topic* mostrado está acortado pero una etiqueta temporal aparece con todo el nombre desarrollado cuando nos situamos con el cursor sobre el *topic*.

Paper. Se trata de la ocurrencia de un *topic*. La clase ocurrencia es mostrada como nombre de un nodo. La herramienta punta muestra la dirección de la fuente enlazada. Si se pincha con el ratón derecho del ratón se abre un cuadro de diálogo para abrir la visualización de la fuente.

Abstract. Es la propiedad de un *topic*. La propiedad clase es mostrada como un nodo nombre. La flecha muestra el valor de la propiedad.

Se puede volver a centrar el árbol estrella en otro *topic* por tipificación de su nombre dentro del campo “vuelta al *topic*” y presionando en la tecla de regreso.

También es posible comenzar una búsqueda desde el árbol estrella. Se pincha con el botón derecho del ratón sobre un nodo *topic* y se accede al comando “Consulta para el nodo” para buscar por ocurrencias del nombre *topic* en las comunicaciones. Para buscar por ocurrencias de múltiples nombres *topic* poniendo los *topics* en el cuadro de búsqueda presionando Shift-Control y pinchando en el nodo simultáneamente. Para lograr la consulta multinodal se pincha en el botón “Consulta por nodos”.

5.8.3.2. Modelo de visualización en índice.

Un caso representativo de visualización de *topic* maps es el proporcionado por Ontopia Navigator. Este navegador ha sido creado expresamente para la navegación en mapas conceptuales contruídos con una herramienta proporcionada por la empresa Ontopia llamada Omnigator. Se trata de una herramienta contruída en soporte Java que permite la construcción y navegación por mapas conceptuales utilizando el lenguaje XTM. La página web de esta empresa, con el objeto de promocionar su software presenta en la página de inicio de Omnigator un apartado llamado “Example applications” dos ejemplos de mapas

conceptuales; uno de ellos “scripts and languages” proporciona una visualización de la relación entre las lenguas y las escrituras con una sintaxis predefinida de relaciones asociativas entre los seis índices interrelacionados. En la página de presentación se presenta en la misma pantalla el listado con los seis índices incluidos y a su derecha un organigrama presentando cada uno de estos índices y el tipo de relación establecida entre los mismos.

Sin embargo, el caso específico que voy a comentar es el siguiente ejemplo. Se trata de un navegador temático construido con la misma herramienta Omnigator pero referido a las Óperas. La visualización es exactamente igual a la que tenemos en un portal de páginas HTML como veremos a continuación en el ejemplo de Operabase presentado como referente en la propia presentación del portal Ópera. Incluso tienen los mismos defectos de visualización como la necesidad de realizar scroll para poder visualizar toda la página, aunque en general podemos decir que es una página web bien construida, con los contenidos temáticos bien estructurados y la presencia de espacios en blanco⁷⁵² que ayudan agrupar los distintos apartados tratados. Por otra parte se trata en ambos casos de unos portales bastante interactivos en el sentido que el usuario no trata con información sino con experiencias⁷⁵³.

La página principal es una página informativa en la que se nos dice cuáles son los objetivos del sitio web, los elementos que constituyen la página y el navegador empleado, Omnigator, con sus características. A la derecha del texto se hace una presentación dinámica y cambiante aleatoriamente de una ópera concreta de la cual se presentan datos como el título, autor de la ópera y autor del libretto.

Lo característico del *topic map* aparece representado a la izquierda. En primer lugar debemos mencionar que los elementos que aparecen en el margen izquierdo no cambian nunca a medida que hacemos la navegación a través de las sucesivas webs. Es decir, en todo momento tenemos ante nuestros ojos el índice por el cual nos orientamos durante la navegación. Haciendo una analogía con el lector de un libro, sería como si éste tuviera presente durante la lectura el índice del libro que está leyendo de forma que no sólo sabe en

⁷⁵² NIELSEN, J. *Usabilidad. Diseño de sitios web*. Madrid: Pearson Educación, 2001.

⁷⁵³ LAUREL, B. *The Art of Human Computer Interface Design*. Reading: Addison-Wesley, 1999.

todo momento el contexto de lo que está leyendo sino que además es capaz por sí mismo de construir lo que se expone en el libro⁷⁵⁴.



Figura 64. Página principal del *topic map* sobre Ópera.

5.8.3.3. Tendencias de visualización con *Topic maps*.

En los últimos estudios sobre técnicas de visualización de contenidos en Web⁷⁵⁵, se está trabajando sobre grafos que sean capaces de representar diagramas nodo-enlace en espacios hiperbólicos en 3D. Estos permiten trabajar con estructuras de representación mayores dado que ya no se están representando datos constreñidos a una estructura jerárquica sino que las estructuras asociativas pasan a ser las preponderantes, especialmente en nuestro modelo de gestión de la información, los *topic maps*. Pero es que no sólo es una cuestión de visualización de la información, sino que cambia la métrica, la forma en que medimos la

⁷⁵⁴ PEPPER, S. *The Italian Opera*. En: <http://www.ontopia.net/operamap/index.jsp>. Consultado 02/06/2009.

⁷⁵⁵ MUNZER, T. H3: Laying Out Large Directed Graphs in 3D Hyperbolic Space. En *Proceedings of the 1997 IEEE Symposium on Information Visualization*, 1997. pp 2-10.

distancia entre los conceptos⁷⁵⁶. El incremento de la distancia entre dos nodos en un espacio euclídeo crece de forma geométrica mientras que en un espacio hiperbólico la distancia tiene un crecimiento exponencial ¿Qué significa esto en términos de navegación en nuestro dominio representado por un *topic map*? Significa que en nuestro limitado espacio físico de un Interfaz Gráfico de Usuario (GUI), podemos visualizar una contextualización en forma de árbol cónico en el sentido de que los *topics* asociados al *topic* central de interés para el usuario se distribuyen en una sección de la esfera en torno a la boca del cono en vez de situarse a lo largo de todo el perímetro esférico. Esto tiene además consecuencias a la hora de introducir conceptos o *topics* ya que en una estructura jerárquica del estilo de una Taxonomía visualizada en árbol, en geometría euclídea, el número de elementos crece de forma exponencial mientras que en un *topic map*, visualizado como el área de una esfera, el número de elementos introducidos crece de forma polinómica ya que las variables de los *topic types* (variedad de *topics*, de acuerdo con la relación asociativa) son más diversas. Esto permite resolver el problema clásico de la colisión entre *topics* análogos mediante la asignación de un menor espacio a los nodos más estrechamente ligados con la consulta del usuario. Cuando retrotraemos nuestra visualización para ver todo el árbol, los únicos nodos que podemos ver en detalle son aquellos que rodean el *topic* raíz, pero si queremos examinar más exhaustivamente los nodos llegamos a perder el contexto⁷⁵⁷.

⁷⁵⁶ GODEHARDT, E. ; BHATTI, N. Using *topic maps* for visually exploring various data sources in a web-based environment. En MAITCHER, L. ; GARSHOL, L. *TMRA 2007*. Berlin : Springer-Verlag, 2008. pp.52-53. En: www.editlib.org/index.cfm/files/paper_11876. Consultado el 15/02/2010.

⁷⁵⁷ MUNZNER, T. H3: Laying out large directed graphs in 3D hyperbolic space. En *Proceedings of the IEEE Symposium on information visualization*. En: <http://graphics.stanford.edu/papers/h3/fig/nab0.gif> Consultado el 15/12/2009. En este caso es una visualización aplicada a las fuentes de información u *occurrences*, pero planteamos esta visualización para *Topic maps* donde los nodos ya no son *occurrences* sino *topics*.

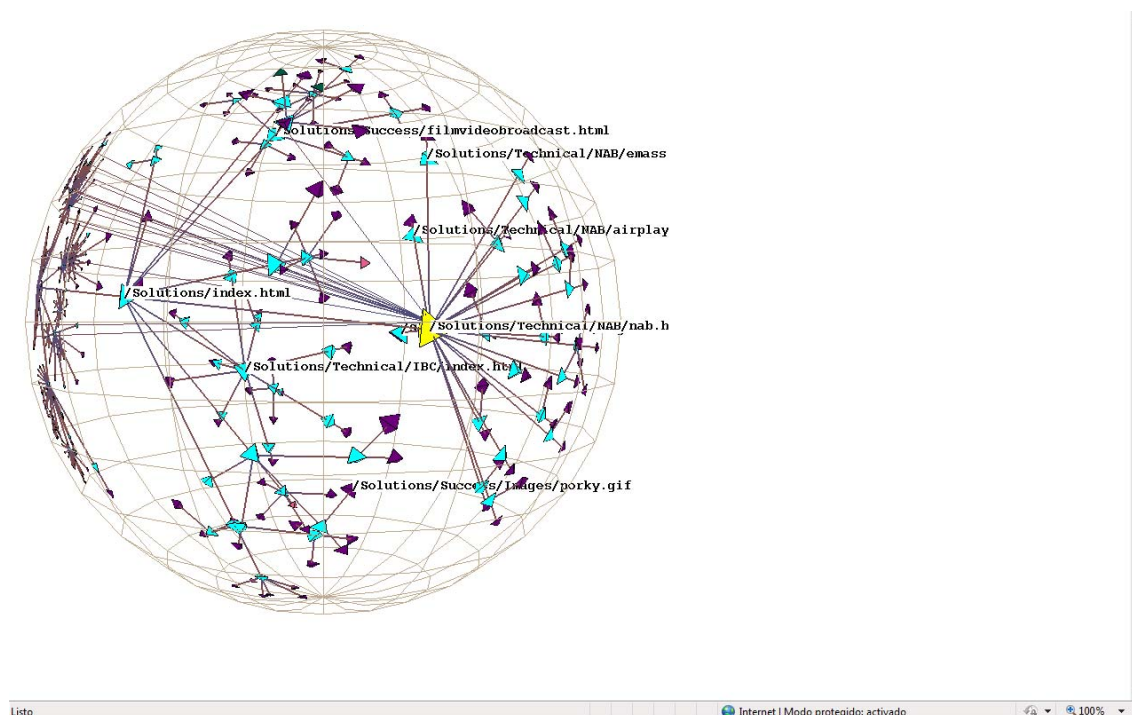


Figura 65. Representación que muestra el modelo propuesto.

Otra novedad que introduce es la relativa a la navegación ya que el GUI deja de ser estático para moverse con el objeto de centrar el *topic* de interés y los *topics* asociados en su entorno siguiendo el modelo de árbol cónico. Por tanto, la navegación hiperbólica permite una visualización del Foco al Contexto de una particular estructura de información con un mínimo desorden o desestructuración del contexto del *topic* de nuestro interés.

5.9. Aplicación de los *Topic maps* al ámbito jurídico.

5.9.1. Representación esquemática de un topic map en el ámbito jurídico.

Cuando se utiliza el lenguaje XML en una organización del marco jurídico se hace por dos de los siguientes objetivos: i) bien para la organización de sus documentos o bien, ii) para hacer que sus aplicaciones interactúen y se comuniquen con las aplicaciones de otros organismos del mismo campo⁷⁵⁸. Con la introducción de XML, la generación y procesamiento de documentación se vuelve más controlable y puede ser automatizada en mayor grado que antes, ya que las aplicaciones pueden comunicarse internamente y externamente a su organización. Sin embargo, lo que buscamos es que recoja los conceptos clave en la documentación de la organización y los disponga de manera conjunta. Y es aquí,

⁷⁵⁸ NOGALES FLORES, T. ; MARTÍN GALÁN, B. *Difusión y recuperación en la Web de documentos jurisprudenciales haciendo uso de la tecnología XML*. En: http://www.bibliociencias.cu/gsd/collect/eventos/index/assoc/HASH0159_dir/doc.pdf. Consultado el 10/02/2010.

donde entran los *topic maps*. Con ellos⁷⁵⁹, se crea un índice por encima de la colección, como podemos ver en la siguiente figura.

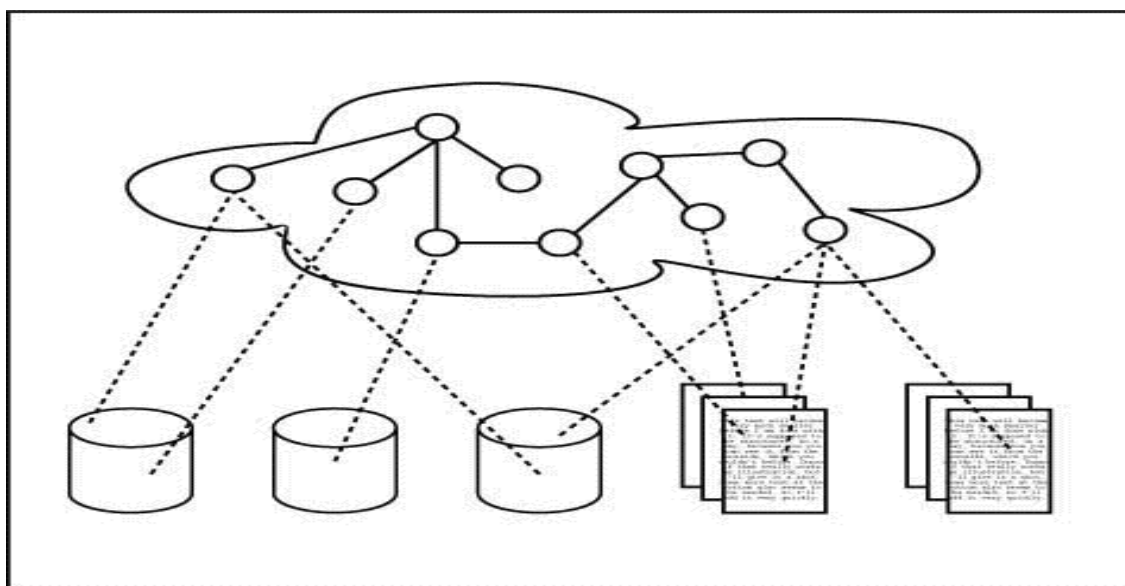


Figura 66. Representación de los topic maps en relación con la colección.

El *topic map*, la nube superior⁷⁶⁰, es un índice que recoge las materias de los documentos y bases de datos jurídicas como pueden ser WestLaw, y con los que se relaciona a través de las líneas o URL's. Esto permite hablar de un contexto informativo con el que los ordenadores pueden realizar una recuperación de información más contextualizada. Es lo que se llama recuperación semántica de información, pero sería más adecuado denominarlo recuperación contextualizada de información⁷⁶¹.

Los *topic maps* toman los conceptos clave de las bases de datos y documentos jurídicos y los relacionan entre sí independientemente de lo que se diga sobre ellos en la información que está siendo indizada⁷⁶²; es decir, cuando en un documento se dice: “El procedimiento de resolución para el caso X consta de los siguientes pasos...”. El *topic map* lo interpreta⁷⁶³:

“La resolución de X es del tipo Q y está contenido en parte de las leyes Y y Z y su procedimiento de mantenimiento reside en el documento W”.

⁷⁵⁹ GARSHOL, L. *What are Topic maps*, 2002. En: http://www.xml.com/lpt/a/2002/09/11/topic_maps.html. Consultado el 10/02/2010.

⁷⁶⁰ Id.

⁷⁶¹ PASSIN, T. *Explorer's guide to the semantic web*. Greenwich : Manning, 2004. pp.66ss.

⁷⁶² DACONTA, M. ; OBRST, L. ; SMITH, K. *The semantic web. A guide to the future of XML, web services, and knowledge management*. Indianapolis : Wiley Publishing, 2003. p.167.

⁷⁶³ AHMED, K. [et al.]. *Professional XML Metadata*. Birmingham : Wrox Press, 2001. p.252.

Por decirlo de otra manera, pasamos a organizar los significados de la información más que la información en sí misma⁷⁶⁴. El resultado es una estructura de conocimiento que rompe el estrecho y rígido marco jerárquico de la actual organización de la información jurídica para explicitar todos sus potenciales contenidos. Esto, hace que esta sea mucho más fácil de buscar porque el usuario deja de buscar como los diseñadores de la información esperan que lo haga. Ello se debe a que existen múltiples maneras de navegación redundante que nos permite obtener la misma respuesta, incluso se pueden emplear las búsquedas en motores para encontrar un buen punto de partida para la navegación. Esto hace que el empleo más común de topic maps sea para la edición de sitios web, dado que facilita el empleo de búsquedas contextuales de información. El *topic map* proporciona la estructura del sitio, y el contenido de las páginas se toma de sus *occurrences*. Asimismo, representa el conocimiento acerca de lo que describe, razón por la cual constituye una herramienta propia de la organización del conocimiento. Pero esto no significa que pueda ser únicamente empleado para ello, sino que también lo puede ser para la organización de contenidos en sistemas gestores, integrando información procedente de distintas fuentes⁷⁶⁵.

La base de los *topic maps* son los *topics*, que representan las materias sobre las que tratan. Por ejemplo, en un *topic map* sobre “Historia del Derecho” uno puede esperar encontrar *topics* tales como “derecho consuetudinario”, “derecho romano”, “potestas”, o “auctoritas”.

Junto con los *topics*, contamos con las relaciones entre los mismos, denominadas *associations*. Estas están tipificadas, lo cual significa la existencia de un abanico amplio de relaciones. Así por ejemplo, la declaración de la relación entre el *topic* poder con los *topics* potestas y auctoritas es una *association* de tipo jerárquico que podemos denominar “manifestado como”.

Las *associations* tienen una característica inusual, y es que cada *topic* empleado en la misma se dice que juega una función, la cual es definida por su tipo de función. Por lo tanto, en la *association* “manifestado como” el *topic* poder juega la función de “fuente”, en tanto que el *topic* auctoritas juega la función de manifestación. Esto significa la posibilidad de estructuración del conocimiento con afirmaciones tales como: “Poder manifestado como auctoritas” o “la auctoritas como manifestación del poder”, lo cual constituye en este caso

⁷⁶⁴ DACONTA, M. ; OBRST, L. ; SMITH, K. Op. cit., p.168.

⁷⁶⁵ SIGEL, A. Towards knowledge organization with Topic Maps. En *XML Europe 2000*. En: <http://www.gca.org/papers/xml europe2000/papers/s22-02.html>. Consultado el 10/02/2010.

una *association* de simetría entre *topics*, para la afirmación de una misma idea en el *topic map*. Sin embargo, la *association* permite declarar distintos tipos de relación entre *topics* de manera complementaria, siempre que las *associations* no estén restringidas a dos de ellos. Esto supone la posibilidad de descomponer una *association* compleja entre varios *topics* como un conjunto de relaciones sencillas entre pares.

La última de las características principales de los *topic maps* es la referida a las *occurrences*. Estas, son fuentes de información relevantes para un *topic*. Para Pepper, las *occurrences* pueden ser su página principal, una foto, un CV, o cualquier otro documento digital⁷⁶⁶. Pueden también estar tipificadas de acuerdo con cada tipo de fuente. Esto significa que si un usuario tiene interés por un *topic* y requiere más información acerca del mismo, este no sólo accede a una colección de enlaces sino que además conoce la pertinencia de cada enlace.

Entonces, ¿cómo se relacionan todos los elementos básicos de un *topic map* del ámbito jurídico? Veamos un ejemplo. En un *topic map* acerca de la ley de procedimiento administrativo común, LPAC, el documento se representa por el *topic* LPAC. Si un jurista quisiera incluir información acerca de las instituciones que han redactado y sancionado la norma tendría que añadir *topics* tales como Cortes Generales, que a su vez se compone de los *topics* Congreso de los Diputados y Senado. Asimismo, tendrá que contar con un *topic* que representa la institución superior sancionadora de las leyes aprobadas en Cortes.

Jefatura del Estado
LPAC
Cortes Generales
Congreso de los Diputados
Senado

Tabla 8. *Topics*

⁷⁶⁶ PEPPER, S. *The TAO of Topic Maps. Finding the way in the age of infoglut*. En: <http://www.ontopia.net/topicmaps/materials/tao.html>. Consultado el 15/02/2010.

Ahora el problema sería cómo establecer la *association* entre todos estos *topics*. Es aquí donde los *topic maps* rompen con el límite de la bidimensionalidad de la información a través de la definición de una tipología de *associations*. Ninguno de los métodos tradicionales de modelado de datos, con la excepción del modelo entidad-relación, ni método de modelado semántico convencional para la recuperación de información tales como taxonomías y tesauros han empleado la tipificación de *associations* de forma significativa.

La *association* entre el *topic* LPAC y el *topic* Jefatura del Estado se puede denominar sanción. Análogamente, la *association* entre LPAC y Cortes Generales puede denominarse autoría, y la *association* entre Cortes Generales y Congreso de los Diputados puede denominarse partitiva dado que el Congreso es una parte de las Cortes.

Es una
Sanción
Autoría
Partitiva

Tabla 9. *Associations*.

Los términos para la designación de las *associations* son escogidos de forma deliberada para poder denotar un sentido, dado que en el modelo básico XTM las *associations* son unitarias, una única etiqueta basta para designar el punto de vista entre ambos *topics*.

No obstante, XTM proporciona un mecanismo para la distinción del sentido de cada *topic* individual por medio del *role* del *topic*. De cada *topic* implicado en una *association* se puede decir que realiza un *role*; así para el presente ejemplo, en la *association* “autoría”, el *topic* “Congreso de los Diputados” realiza el role de autoría y el *topic* LPAC realiza el rol de “ley”. Análogamente, en la *association* “partitiva” el *topic* “Cortes Generales” tiene el *role* de genérico, en tanto que “Congreso de los Diputados” tiene el *role* de “específico”.

En algunos contextos puede ser preferible escoger distintos términos. Así, en nuestro caso, la *association* puede ser denominada subsidiariedad cuando “Cortes Generales” realiza el *role* de supraorganismo con “Congreso de los Diputados” jugando el *role* de suborganismo. Esta estructura indefinida pero muy flexible nos permite seguir actualizando el *topic map* con otros *topics* como pueden ser “ley orgánica”, que denota un tipo de ley diferenciado del resto de leyes.

Para hacer esto, se podría definir la *association* “es una” existente entre el *topic* “ley orgánica” y el *topic* “LPAC”, de forma que toda la relación se leería como la “LPAC” es una “ley orgánica”. Podríamos caracterizar el *role* del *topic* “ley orgánica” como “ley aprobada por una mayoría absoluta” y el *role* de “LPAC” como “instancia”. Así la *association* entre el *topic* “Jefatura del Estado” y el *topic* “ley orgánica” como “sanciona”, con el *topic* “Jefatura del Estado” ejerciendo el *role* de “sancionador” y el *topic* de “ley orgánica” ejerciendo el *role* de “ley aprobada por una mayoría absoluta”.

Ley aprobada por una mayoría absoluta
Instancia
Sancionador
Supraorganismo
Suborganismo
Autor

Tabla 10. *Roles.*

La representación esquemática con un mapa conceptual de la parte del *topic map* correspondiente a la LPAC, con su correspondientes *topics*, *occurrences*, *associations* y *roles* lo vemos a continuación.

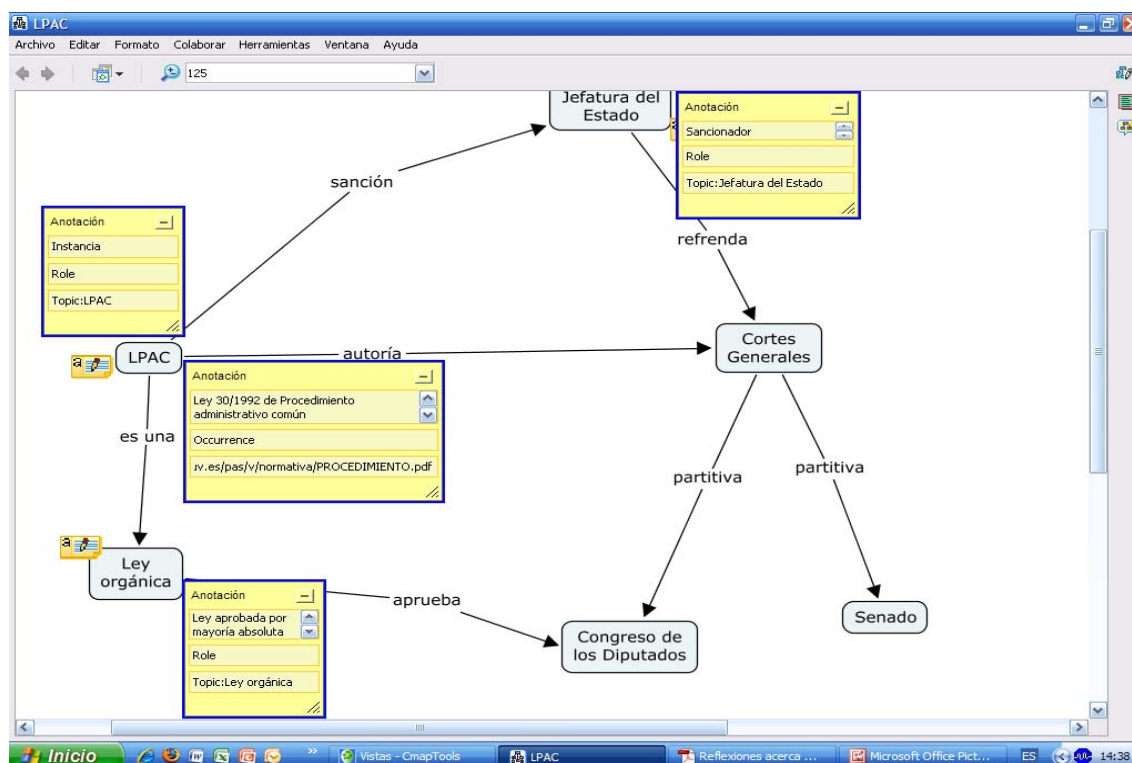


Figura 67. Representación gráfica de un topic map sobre LPAC.

Este mecanismo de *topic role<association>topic role*, proporciona un amplio abanico de posibilidades para la realización de búsquedas y la navegación precisa, dado que cualquiera de sus elementos puede ser empleado para ambos procesos; no obstante, la construcción inicial del *topic map* puede suponer un notable esfuerzo de análisis semántico y lingüístico para la elección de los términos pertinentes para cada uno de los elementos del *topic map*.

Los *topics* tienen otras características además de los *roles* y *associations*, son las *occurrences*. Estas son recursos de información discretos con una relación con un *topic*, como por ejemplo el texto de la LPAC en formato pdf. En el esquema del *topic map* se muestra un hipervínculo a una localización en la web donde se puede recuperar el documento en pdf. Las *occurrences* pueden estar en cualquier formato. Recursos digitales como archivos de texto, páginas HTML, documentos XML y bases de datos son accesibles a través de un hipervínculo insertado en el *topic map*. Los recursos informativos no digitales son accesibles a través de las referencias bibliográficas convencionales.

Además de estos cuatro conceptos clave, la sintaxis XTM incluye dos características más de los *topics*: el identificador público de materia, y el *scope*. Los identificadores públicos de materia, PSI, son definiciones formales de una materia que la identifican de forma unívoca para todos los *topic maps*, lo que permite emplear distintos términos para designar la misma materia. La forma de definir una materia con sus conceptos inmediatos subordinados y supraordenados constituye una taxonomía.

El *scope* es empleado para indicar el contexto específico en el cual es válida una *association*. Tanto el PSI como el *scope* son igualmente útiles para el usuario y los ordenadores. Para el usuario, un navegador *topic map* tendría un interfaz capaz de mostrar una definición que indicase la identidad exacta de un *topic*. Para los ordenadores, el PSI y el *scope*, pueden ser empleados para la fusión de *topic maps* a fin de que la máquina sea capaz de conocer a partir del PSI qué *topics* con denominaciones diferentes se refieren a la misma materia, y en qué *scope* debe ser empleado cada sinónimo. Así por ejemplo, un *topic map* puede incluir un *topic* indicando que un recurso tiene contenidos que se relacionan con la lengua Ladina. La inclusión en el *topic map* del PSI pertinente identifica de forma desambigua que la lengua a la que se refiere es un dialecto del español hablado por los judíos sefarditas; y se encuentra representado en ISO 639 por el código “lad”⁷⁶⁷. Por otro lado, un recurso de información en francés que estuviera incluido por el mismo *topic map*, emplearía el nombre en francés (judéo-espagnol) para esta misma lengua. Las *associations* que empleen este término en francés y aquellas que empleen su equivalente en inglés se muestran como referidas a la misma lengua por medio de hiperenlaces al mismo PSI. La separación de los *scopes* en cada una de las *associations* indica que uno es el término en inglés y el otro el término en francés. El empleo del *scope*, *topic* y *association names* puede ser duplicado en distintas lenguas, permitiendo que el usuario pueda navegar por las distintas traducciones de un mismo *topic map*.

⁷⁶⁷ ISO 639-2 de los registros de autoridad de la Library of Congress. En: <http://www.loc.gov/standards/iso639-2/langhome.html>. Consultado el 10/02/2010.

5.9.2. Casos prácticos en la aplicación de *Topic maps* en el ámbito jurídico.

Existen ya algunas experiencias en el diseño de la arquitectura de información de sitios web del ámbito jurídico con *topic maps*⁷⁶⁸. Una de las experiencias más documentadas y que mayor impacto ha tenido es la realizada por la consultora Coolheads de Newcomb y Biezunski para la agencia tributaria norteamericana⁷⁶⁹. Esta experiencia, muy comentada en congresos de XML y publicaciones sobre *topic maps* nos ha servido de referente para plantear una arquitectura de información con *topic maps* para el Centro de Estudios Políticos y Constitucionales. Es por esta razón que vamos a proporcionar algunas de las características más importantes de esta iniciativa que nos servirán en el siguiente capítulo para justificar la metodología planteada para la edición de *topic maps*.

La Agencia Tributaria⁷⁷⁰ de Estados Unidos genera una gran cantidad de documentación relativa a normativa y legislación sobre recaudación que es importante sea bien difundida tanto entre las personas físicas como entre las organizaciones. Sin embargo, cada una de estas publicaciones es producida e indizada de forma independiente. Cuando se reúne una serie de publicaciones en un único CD-ROM, surge el problema de cómo unificar los índices de forma que éstos puedan proporcionar acceso al contenido de cualquier documento de cada una de las publicaciones independientemente del formato del documento. Así por ejemplo la agencia tributaria estadounidense hizo un CD-ROM en el 2001 sobre todo tipo de impuestos que hace uso de los *Topic maps* para crear una estructura de navegación a través de toda la variedad de publicaciones incluidas en el CD-ROM⁷⁷¹.

⁷⁶⁸ AHMED, K. *Topic maps- A practical introduction with case studies*. En: https://ssl.bnt.com/idealliance/papers/xmle02/dx_xmle02/papers/03-05-01/03-05-01.pdf. Consultado el 15/12/2009.

⁷⁶⁹ COOLHEADS CONSULTING. En: <http://www.coolheads.com/publications.htm>. Consultado el 15/12/2009.

⁷⁷⁰ BIEZUNSKI, M. The story of a *Topic maps* Use Case: The IRS Call Center Tax Map. En *Proceedings of XML Europe 2003*. London: Idealliance, 2003. En: www.coolheads.com/MBPUBS/xmleurope2003/irstm4.ppt. Consultado el 15/12/2009.

⁷⁷¹ TECHQUILA. En: <http://www.techquila.com/papers/images/03-05-01-fig01.jpg>. Consultado el 15/02/2010.

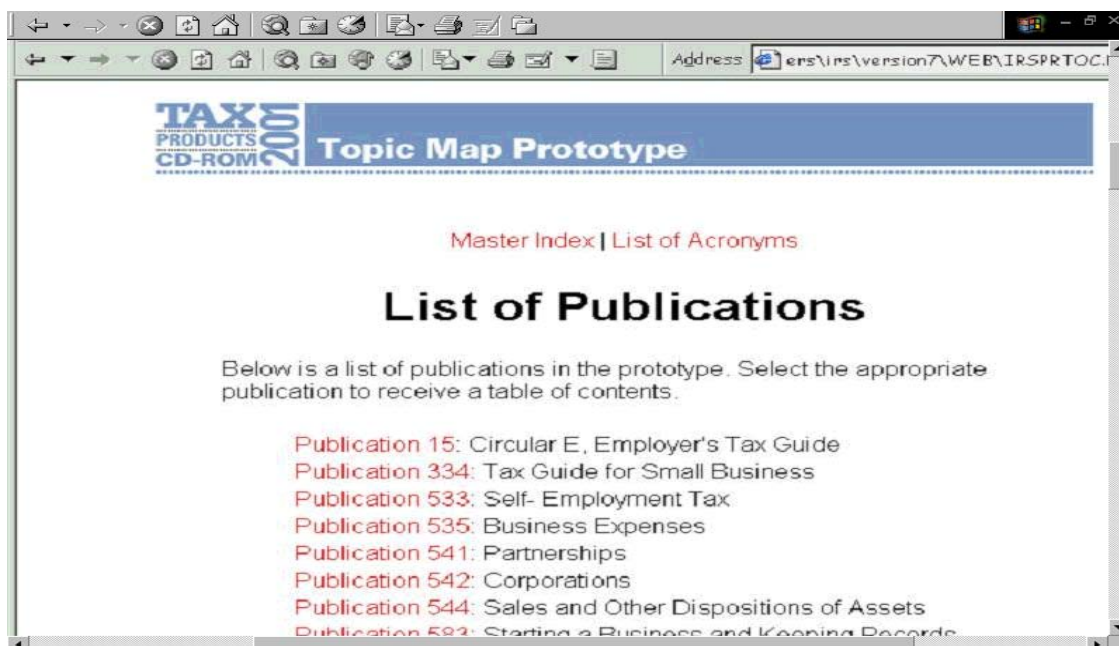


Figura 68. Ejemplo de *topic map* para publicaciones sobre derecho tributario.

El acceso a la colección es a través de una tradicional lista de contenidos en la que cada publicación tiene su propia hoja de contenidos que incluye un enlace a un índice generado a partir de los nombres de los *topics* encontrados en el procesamiento de cada documento. También se proporciona un acceso al índice desde dentro del propio documento a través de enlaces. Estos enlaces se generan cuando se localiza la ocurrencia de un *topic* dentro del contenido de un documento. Los enlaces entonces llevan al usuario a una página generada por el propio *topic*, esta página agrega todas las ocurrencias del *topic* que hay en el CD-ROM, proporcionando al usuario un rápido acceso a toda la variedad de información relacionada con los términos del índice. Así por ejemplo podemos navegar saltando desde el concepto de “Impuestos estimados” a información sobre el concepto de “pequeños negocios” o el “autoempleo” a través de los títulos publicados y los fragmentos de contenido mostrados en el índice ya que éstos permiten al usuario escoger rápidamente los documentos más relevantes. Por lo tanto, es el empleo de los enlaces a ocurrencias de los términos del Índice lo que proporciona una gran flexibilidad a la estructura del *topic map* ya que permite que los *topics* del índice nos lleven únicamente a las correspondientes ocurrencias con las que se relaciona cada *topic*. Relaciones adicionales a otros *topics* del índice también pueden incluirse dentro del *topic map* permitiendo tanto una organización jerárquica de los *topics* en el índice como facilitando las referencias cruzadas entre los *topics*.

La conexión entre el *topic map* y las páginas HTML se ha realizado a través de una herramienta denominada *Topic map Loom*. El punto de partida para la generación del *topic map* ha sido una colección de documentos fuente que estaban en SGML. Con la herramienta ya mencionada se hace una extracción automática de *topics* y ocurrencias de los *topics* a partir de esta colección de documentos fuente. El *topic map* generado de esta manera es entonces usado para generar la navegación entre las páginas HTML sobre las que se implementa. En general, los *topic maps* son bastante apropiados para situaciones en las que un corpus de documentos contiene varios índices generados de forma independiente; ya que los *topic maps* han sido diseñados desde el principio para ser fusionables de forma que los *topic maps* individuales para cada publicación por separado puede ser combinada empleando los principios de fusionabilidad del paradigma *topic map* para crear un único *topic map* global.

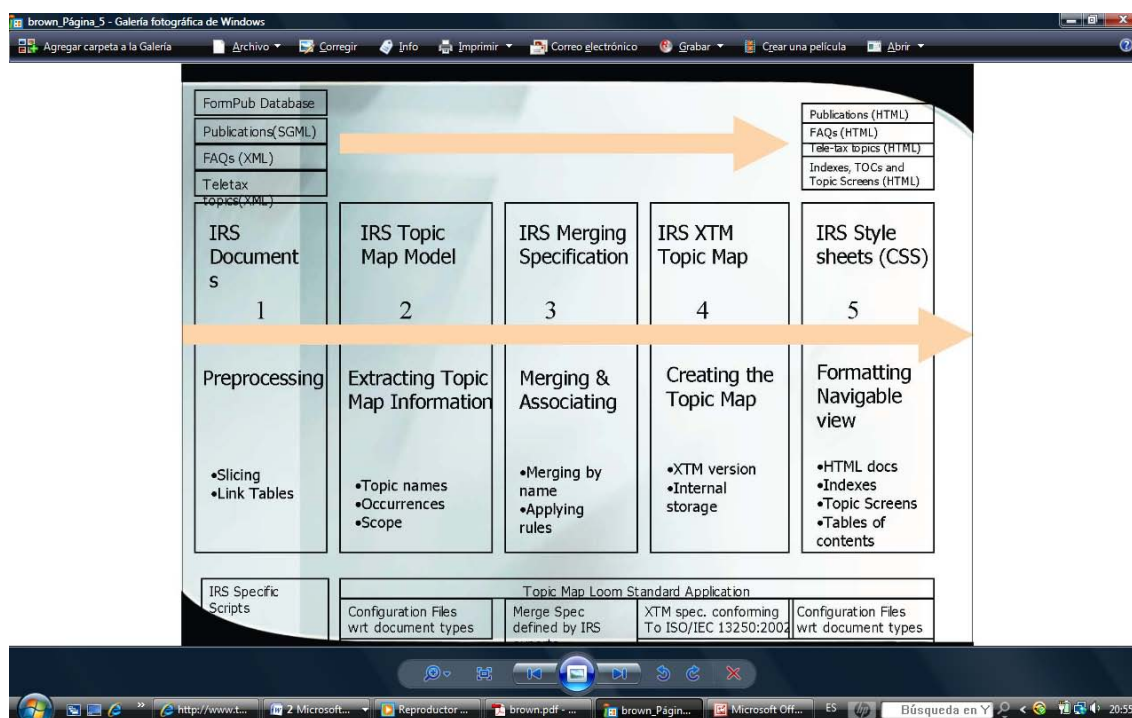


Figura 69. Arquitectura de edición del *topic map* Tax Map.

Actualmente el proyecto IRS se encuentra trabajando en la búsqueda de una tecnología que permita mejorar los índices individuales de cada publicación por separado para convertirlos en una colección de índices apropiada para su publicación en un CD-ROM⁷⁷². Se está buscando un modo de representar los *topics* y sus *occurrences* sino también introducir

⁷⁷² BROWN, D. ; BIEZUNSKI, M. *IRS Tax Map. Electronic research tool for tax law assistants*. En: http://ipgems.com/present/06_Brown-Biezunski_10-03.pdf. Consultado el 15/02/2010.

relaciones más complejas entre los *topics* para ayudar a guiar al usuario entre la documentación. Ambas necesidades deben encontrar una solución basada en la representación de estas estructuras con un *topic map*. El empleo del *topic map* también proporciona a los editores la flexibilidad necesaria para mejorar la navegación con referencias cruzadas y referencias externas sin la necesidad de modificar la tecnología subyacente y los procesos empleados para generar esa navegación.

Hasta el momento, las pruebas de usabilidad de la aplicación IRS Tax Map han mostrado a los evaluadores que los usuarios más entusiastas han sido los consultores del *call centre* de la propia agencia dedicados a trabajar en la resolución de dudas planteadas por los contribuyentes sobre cómo pagar impuestos, así como aquellos con menos experiencia en materias muy específicas sobre las que tienen menos experiencia.

De la evaluación se constata que no han tenido problemas respecto a cómo navegar por el *topic map* ya que todos han sido capaces de emplearlo inmediatamente. Algunos consultores prefieren navegar entre los documentos, otros prefieren usar los índices para encontrar un *topic* y otros prefieren emplear el motor de búsqueda⁷⁷³. En ocasiones, algunos consultores han confesado que ante algunas preguntas no habrían sabido proporcionar la respuesta más oportuna de no haber sido por el Tax Map. Con lo cual, el empleo del Tax Map proporciona más confianza a los consultores y una mejora en la calidad de las respuestas al público.

A raíz del éxito de esta herramienta⁷⁷⁴, la propia agencia tributaria comenzó a distribuir en un CD-ROM el programa entre aquellos contribuyentes que así lo solicitaran, para ayudarles directamente en la resolución de dudas acerca de su declaración de la renta sin necesidad de comunicarse con un *call centre*. Sin embargo, hubo tantas solicitudes que tres años más tarde se hizo universal el acceso al Tax Map directamente por Internet a través del sitio web del IRS Tax Map⁷⁷⁵.

⁷⁷³ Brown, D. ; Biezunski, M. *IRS Tax Map. Electronic Research tool for Tax Law Assistors*. En: http://www.ipgems.com/present/06_Brown-Biezunski_10-03.pdf. Consultado el 02/02/2010.

⁷⁷⁴ *IRS Tax Map*. En: <http://taxmap.ntis.gov/taxmap/>. Consultado el 02/02/2010.

⁷⁷⁵ *Id.*

Topic Maps y Tecnologías de la Información Jurídica: un modelo de recuperación de información para bibliotecas digitales semánticas jurídicas.

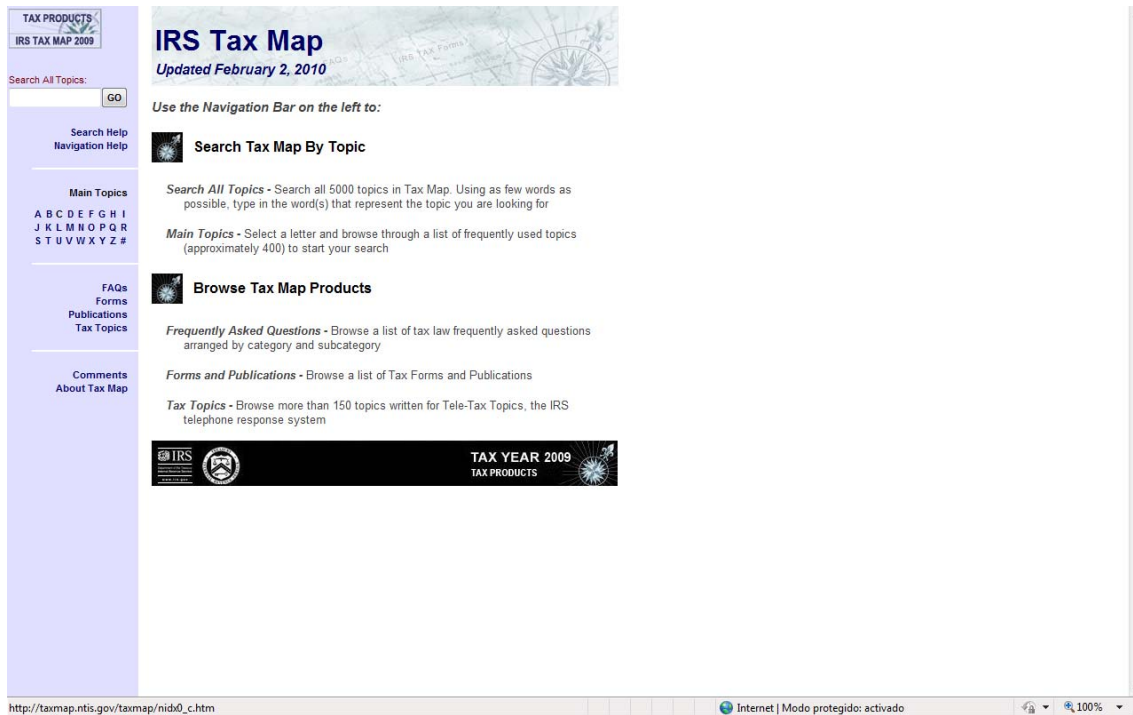
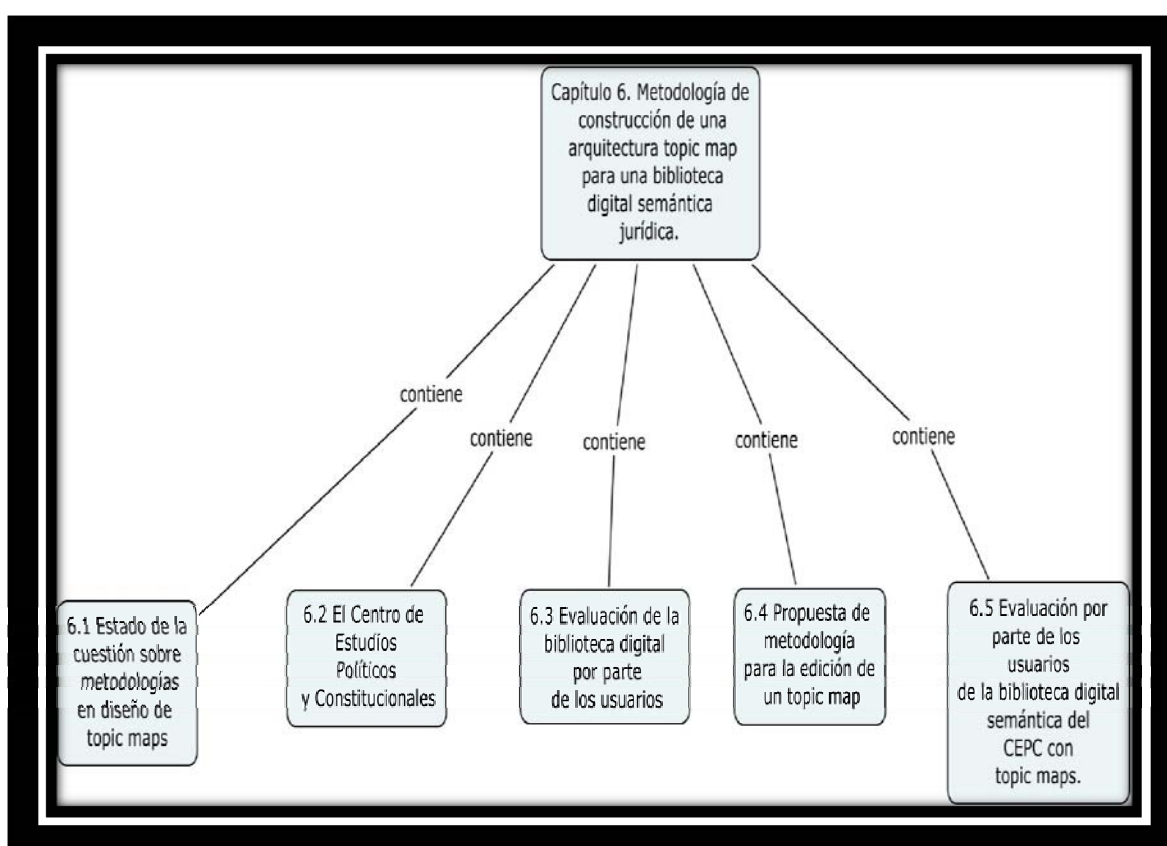


Figura 70. Ejemplo del sitio web de acceso al *topic map* para el ciudadano desde Internet.

BLOQUE II. Maqueta de aplicación del modelo Topic map a una biblioteca digital semántica jurídica. El portal del CEPC.

CAPÍTULO 6.

Metodología de construcción de una arquitectura *Topic map* para una biblioteca digital semántica jurídica.



Introducción.

El propósito del capítulo es explicar la propuesta de una metodología planteada para el *topic map* de la biblioteca digital semántica del CEPC en el contexto de la evaluación que sus usuarios realizan por un lado de los servicios de la biblioteca, y por otro del *topic map* construido a partir de sus sugerencias y análisis cualitativo de sus habilidades para la localización de información.

Esta metodología, la hemos concretado, tal y como sugerimos durante el desarrollo de la presente tesis, con la inserción de ejemplos del *topic map* diseñado para el portal del Centro de Estudios Políticos y Constitucionales. La hemos ido puliendo con las entrevistas realizadas entre los distintos sujetos que participaron en el diseño y uso del *topic map*. Está diseñada explícitamente para un desarrollo inicial que dote de estructura a la biblioteca digital semántica del CEPC. Sin embargo, puede ser adaptada a otros campos, tales como web social o gestión del conocimiento corporativo.

6.1. Estado de la cuestión sobre metodologías en diseño de *topic maps*.

Ahmed⁷⁷⁶, Pepper⁷⁷⁷ y Garshol⁷⁷⁸ nos indican que todo proyecto *topic map* necesita contar con una ontología, y la mayoría de estos necesitan desarrollar una propia. Para los planteados hasta ahora, es ésta una de sus mayores dificultades, a causa del desconocimiento sobre cómo abordarlo ante la carencia de publicaciones sobre el diseño de ontologías *topic map*. Esto impide la existencia de una corriente de retroalimentación que progresivamente mejore, a partir de la evaluación, las técnicas de diseño.

Lo primero que debemos decir es que no existe una sola metodología “correcta” para desarrollar *topic maps*. Para Garshol⁷⁷⁹, el diseño de *topic maps* supone un proceso previo de diseño de una ontología adecuada por igual a las demandas informativas de sus usuarios y a la organización del conocimiento que se desea conseguir para un dominio concreto de conocimiento. De hecho, asistimos a un seminario⁷⁸⁰ acerca del diseño de *topic maps* impartido por Garshol. En el mismo partía del concepto de ontología y demostraba que el diseño de un *topic map* suponía hablar del diseño de una ontología, pero con sus características peculiares. Por esta razón, partimos de los autores más destacados en

⁷⁷⁶ AHMED, K. *Topic map patterns for information architecture*. En: www.techquila.com/tmsinia.html. Consultado el 15/02/2010.

⁷⁷⁷ PEPPER, S. *Creating a topic map*. En: www.topicmaps.com/tm2008/pepper3.pdf. Consultado el 15/02/2010.

⁷⁷⁸ GARSHOL, L. *A citizen's portal for the city of Bergen*.

En: http://www.informatik.uni-leipzig.de/~tmra/2007/slides/garshol_TMRA2007_bergen.pdf. Consultado el 15/02/2010.

⁷⁷⁹ GARSHOL, L. *Modelling with Topic maps*. En:

http://www.topicmaps.com/tmc/presentation.jsp?conf=TM2007&id=Lars_Marius_Garshol-TM2007-Modelling_with_Topic_Maps
Consultado el 15/02/2010.

⁷⁸⁰ *The First International Topic Maps Users Conference*. Oslo, 2007. En:

<http://www.topicmaps.com/tmc/conference.jsp?conf=TM2007>. Consultado el 15/02/2010.

ontologías, como el trabajo realizado por Guarino⁷⁸¹, para plantear el diseño de nuestro *topic map*.

A finales de los años 90 y principios del siglo actual hemos sido testigos del creciente interés por parte de distintos campos de conocimiento por construir ontologías y su posterior reutilización en el diseño de otras nuevas⁷⁸². Se comenzó a contar con el empleo de métodos semiautomáticos capaces de disminuir el cuello de botella que se estaba produciendo en la adquisición de conocimiento por medio del empleo de metodologías de desarrollo de ontologías. Hasta aquel entonces cada grupo de trabajo seguía sus propios principios, criterios de diseño y metodologías para la construcción manual de ontologías. La ausencia de directrices comunes retrasaba el desarrollo de ontologías que pudieran ser interoperables, lo que dificultaba su empleo para su reutilización y uso en aplicaciones finales.

En 1996, tuvo lugar el primer seminario acerca de ingeniería ontológica en el seno del décimo segundo congreso europeo de inteligencia artificial⁷⁸³. Su objetivo era explorar un conjunto de principios, criterios de diseño y reglas de buenas prácticas que pudieran ser aprovechadas por cualquier diseñador de ontologías. Un segundo seminario tuvo lugar al año siguiente con el mismo objetivo pero esta vez se celebró en el ámbito de los simposios del AAAI en Stanford. Uno de los principales aspectos tratados aquí era precisamente el empleo de ontologías para el diseño y evaluación de ontologías. Desde entonces han sido presentados numerosos aspectos relacionados con la metodología a emplear en el proceso de desarrollo de las ontologías en los congresos y seminarios internacionales del campo de la ingeniería ontológica como los congresos EKAW⁷⁸⁴.

Para comprender lo que significa la definición de una metodología partimos de la concepción que nos proporciona el IEEE acerca de la misma. Para este organismo, es una serie integrada y comprensiva de técnicas o métodos para la creación de una teoría general

⁷⁸¹ GUARINO, N. Formal ontology in information systems. En: Guarino, N. (edit.) *I International Conference on Formal Ontology in Information Systems (FOIS'98)*. Trento : IOS Press, 1998. pp.3-15.

⁷⁸² GÓMEZ-PÉREZ, A. ; FERNÁNDEZ-LÓPEZ, M. ; CORCHO, O. *Ontological Engineering*. Berlin : Springer-Verlag, 2004. p. 107

⁷⁸³ WAHLSTER, W. (ed.). *Proceedings of the 12th European Conference on Artificial Intelligence (ECAI-96)*. Chichester : Wiley, 1996.

⁷⁸⁴ EKAW. *Knowledge Engineering and Knowledge Management by the Masses*. En: <http://ekaw2010.inesc-id.pt/index.html>. Consultado el 20/02/2010.

de sistemas sobre cómo debe ser realizada una clase de procesos de pensamiento⁷⁸⁵. Esto permite que las metodologías sean empleadas en organización y representación del conocimiento⁷⁸⁶; es decir, presentan “qué”, “quién” y “cuando” una actividad dada debe ser realizada. Por ejemplo, Noy y McGuinness⁷⁸⁷ plantean reglas para ayudar a la toma de decisiones durante el proceso de diseño de ontologías, que a grandes rasgos, son también aplicables al diseño de *topic maps*. Reglas que para el caso de los *topic maps* hemos reformulado así:

1. No existe una forma correcta para modelar un dominio, sino que siempre hay alternativas viables. La mejor solución casi siempre depende de la aplicación que se tiene en mente y las extensiones que se anticipen.
2. El desarrollo de *topic maps* es un proceso necesariamente iterativo.
3. Las materias en los *topic maps*, deben ser pertinentes tanto para los documentos digitales como para las relaciones propias del dominio de conocimiento jurídico.

En general, estas reglas lo que permiten es decidir para qué vamos a emplear los *topic maps*, así como su nivel de detalle, teniendo en cuenta distintas alternativas para determinar cuál se adecúa mejor a la tarea proyectada, cuál es la más intuitiva, la más extensible o la más sencilla de mantener. Decisiones, todas ellas, regidas por la idea de que un *topic map* es un modelo de la realidad del cual, una vez definida su versión inicial, podemos evaluar y depurar discutiéndolo tanto con expertos en el área como con usuarios. Este proceso de diseño iterativo se produce durante todo el ciclo de vida del *topic map* para su mantenimiento y actualización.

6.2. El Centro de Estudios Políticos y Constitucionales.

El Centro de Estudios Políticos y Constitucionales es un organismo autónomo dependiente del Ministerio de la Presidencia cuya misión consiste en fomentar el análisis de la realidad

⁷⁸⁵ IEEE Standard Glossary of Software Engineering Terminology. New York : IEEE Computer Society, 1990. En: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=159342>. Consultado el 15/02/2010.

⁷⁸⁶ WIELINGA, B.J., SCHREIBER, A. T., BREUKER, J.A., KADS: A Modelling Approach to Knowledge Engineering. *Knowledge Acquisition*, vol. 4, nº1,1992. p.15.

⁷⁸⁷ NOY, N. ; MCGUINNES, D. *Desarrollo de ontologías-101: Guía práctica para crear tu propia ontología*. En: http://protege.stanford.edu/publications/ontology_development/ontology101-es.pdf. Consultado el 20/02/2010.

jurídica y sociopolítica nacional e internacional, prestando especial atención a todas aquellas cuestiones relativas al Derecho español y a las instituciones de países iberoamericanos, así como a las relaciones de éstos con Europa⁷⁸⁸.

Para cumplir sus objetivos, el Centro promueve estudios sobre los sistemas políticos constitucionales y administrativos nacionales e internacionales, edita publicaciones y mantiene un fondo bibliográfico y documental sobre Derecho Constitucional, Teoría del Estado, Ciencia Política e Historial Política de España. Asimismo, organiza seminarios de enseñanzas especializadas relacionadas con las mencionadas materias, entre las cuales destaca el Curso de Derecho Constitucional y Ciencia Política⁷⁸⁹.

6.2.1. Origen y evolución del CEPC.

El periodo comprendido entre 1975 y 1978 es conocido en la historia política de España como la “transición democrática. Se trataba de una transformación del sistema político al pasar de un régimen dictatorial a un sistema de democracia parlamentaria “en transición”⁷⁹⁰. Esto supuso una transformación controlada y pactada del sistema político, por el cual pasamos del sistema regido por las leyes fundamentales al sistema que se inicia formalmente con las leyes de la Reforma Política⁷⁹¹. Las instituciones que sustentaban el antiguo modelo de Estado tuvieron que evolucionar hacia modelos propios de las democracias parlamentarias. Las antiguas Cortes franquistas debieron dar paso a un sistema bicameral, basado en la existencia de un Senado y un Congreso de Diputados⁷⁹².

El Senado había cerrado sus puertas desde el comienzo de la dictadura de Primo de Rivera en 1923, y el Congreso en 1936. Al término de la Guerra Civil, el Congreso acogió las Cortes Franquistas y el Senado se convirtió en la sede del Consejo Nacional del Movimiento, donde también tenía su sede el Instituto de Estudios Políticos, organismo dedicado a la investigación en ciencias político-sociales y a la formación de mandos políticos.

⁷⁸⁸ REAL DECRETO 1269/1997, de 7 de agosto de 1997, de organización y funcionamiento del Centro de Estudios Políticos y Constitucionales (BOE n°214). En: <http://www.boe.es/boe/dias/1997/08/07/pdfs/A24078-24080.pdf>. Consultado 15/02/2010.

⁷⁸⁹ Id.

⁷⁹⁰ REDERO SAN ROMÁN, M. *La transición a la democracia en España*. Madrid : Marcial Pons, 1996. pp.15ss.

⁷⁹¹ HERNÁNDEZ GIL, A. *El cambio político español y la Constitución*. Barcelona: Planeta, 1982. p.248.

⁷⁹² TEZANOS, J. La crisis del franquismo y la transición democrática. En TEZANOS, J. ; COTARELO, R. ; BLAS, A. de.(eds.) *La transición democrática española*. Madrid: Sistema, 1989. pp.9-30.

En 1977, durante la transición democrática, nace el Centro de Estudios Constitucionales, cuando en octubre se realiza una reforma del Ministerio de Presidencia por medio de un decreto en el que se suprimen varios organismos del ministerio, sobre todo aquéllos que habían estado más próximos al régimen anterior, como es el caso del Instituto de Estudios Políticos⁷⁹³.

Por otra parte, se decreta el cambio de denominación del Instituto de Estudios Administrativos, unidad especializada del Instituto Nacional de Administración Pública, por la denominación de Centro de Estudios Constitucionales. A este nuevo organismo se le adscriben los recursos materiales y humanos del desaparecido Instituto de Estudios Políticos, además de su labor editorial y formativa, pero adaptada al nuevo sistema político.

El Instituto de estudios Políticos nació por Decreto de 9 de septiembre de 1939 como un organismo dependiente de la Junta Política de Falange Española Tradicionalista y de las JONS, lo que le confirió un carácter marcadamente político al servicio del nuevo régimen franquista. Su sede se estableció en el Palacio del Senado que había cerrado sus puertas durante la dictadura de Primo de Rivera en 1923.

Este instituto se creó con dos objetivos principales⁷⁹⁴: i) Investigar con “criterio político y rigor científico los problemas de la vida administrativa, económica, social e internacional de la patria”, y ii) constituirse en escuela de “formación política superior de elementos destacados de las nuevas generaciones”.

⁷⁹³ GÓMEZ DEL PULGAR RODRÍGUEZ DE SEGOVIA, G. *Evaluación de las revistas editadas por el Centro de Estudios Políticos y Constitucionales* [memoria de licenciatura]. Getafe: Universidad Carlos III, 2000. pp.10ss.

⁷⁹⁴ Decreto de 9 de septiembre de 1939. Artº 1º. BOE de 11 de septiembre de 1939. En: <http://www.boe.es/datos/imagenes/BOE/1939/254/A05061.tif>. Consultado el 15/02/2010.

DECRETO DE 9 DE SEPTIEMBRE DE 1939 creando el Instituto de Estudios Políticos dependiente de la Junta Política de F. E. T. y de las J. O. N. S.

La Junta Política, Delegación Permanente del Consejo Nacional de Falange Española Tradicionalista y de las J. O. N. S., está llamada a ser el órgano a través del cual se promueva la reforma del Estado, para que responda en todos sus aspectos a la ambición histórica del Movimiento Nacional.

Por ello, es de gran conveniencia la creación de un organismo que, dependiendo de la Junta, investigue con criterio político y rigor científico los problemas y manifestaciones de la vida administrativa, económica, social e internacional de la Patria. Dicho organismo podrá ser, al mismo tiempo, escuela para la formación política superior de elementos destacados de las nuevas generaciones.

Figura 71. Boletín Oficial del Estado donde se publica la constitución del Instituto de Estudios Políticos.

El cumplimiento de estos objetivos se pretendía lograr por medio de seis tipos de actividades⁷⁹⁵:

- 1) Intervención en la vida política del régimen bien como institución o bien por la participación política de los miembros del instituto.
- 2) Por otro lado el Instituto colaboraba en la tarea legislativa del Estado mediante la elaboración de proyectos de ley y dictámenes.
- 3) Presencia en la sociedad mediante la organización de eventos socioculturales.
- 4) Formación política a través de cursos fundamentalmente relacionados con las concepciones falangistas de la teoría política.
- 5) Fomento de la investigación dado que en este instituto se contaba desde 1923 con la biblioteca del Senado, integrada por diversas colecciones. Si bien los fondos de esta biblioteca terminarían por quedarse en el Senado y los fondos de la actual biblioteca del CEPC son los adquiridos por el Instituto a partir de 1940.
- 6) Actividad editorial donde destacó tanto por sus publicaciones periódicas como por sus colecciones de monografías.

⁷⁹⁵ GÓMEZ DEL PULGAR RODRÍGUEZ DE SEGOVIA, G. Op. cit., p.23.

- Artículo tercero.*—Son fines del Instituto:
- a) El estudio, conforme a un plan trazado, de los problemas fundamentales que corresponden a cada Sección y el acopio de la información necesaria a dicho objeto. El Director del Instituto someterá a la Junta Política los resultados de los trabajos.
 - b) El asesoramiento de la Junta Política, la Secretaría General y los distintos Servicios del Movimiento, en las materias propias de su competencia, cuando fuere solicitado por estos Organismos.
 - c) El dictamen sobre asuntos o proyectos de Gobierno, cuando fuere requerido a darlo por algún Órgano del mismo (con aprobación del Ministro Presidente de la Junta Política), por la Junta o por la Secretaría General del Movimiento.
 - d) La orientación, de acuerdo con las directrices fijadas por la Secretaría General, de la actuación de todas las representaciones del Partido en los Organismos oficiales, en las materias propias del Instituto.
 - e) La orientación de las Instituciones del Movimiento, para la formación de sus Jerarquías.
 - f) La dirección de los trabajos que corporativamente realicen las profesiones liberales para el estudio de los problemas de la vida nacional.

Figura 72. Publicación de los fines del instituto.

Respecto a la organización del Instituto de Estudios Políticos (IEP) no sufrió cambios hasta que se transformó en CEPC. Estaba constituida por una dirección ocupada por cargos de confianza del régimen, quienes dirigían sus actividades, hacían que se redactaran los proyectos y dictámenes, así como la selección de las personas a cargo de la docencia del instituto.

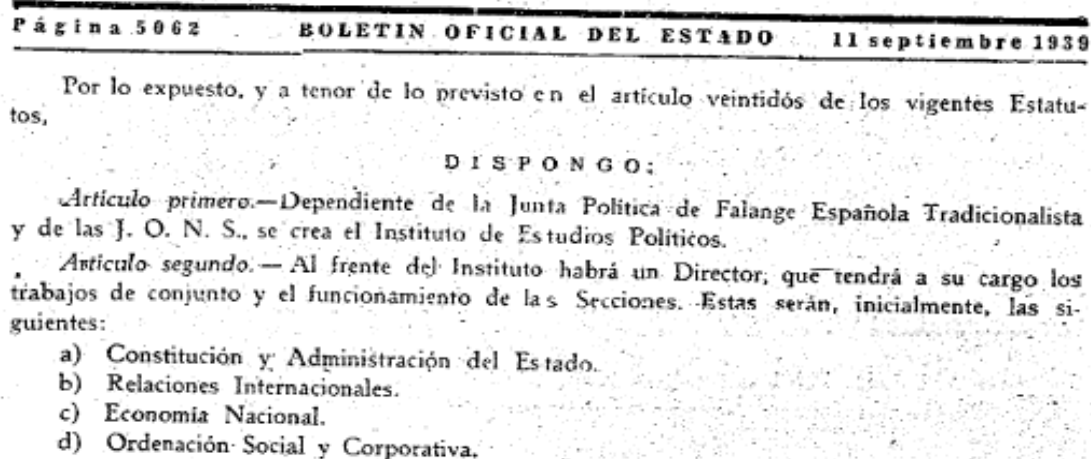


Figura 73. Organización del instituto.

Asimismo, existía una comisión permanente formada por el director, los presidentes de sección, el secretario general técnico, el director de los cursos y algunos procuradores en Cortes. Tenían por función la programación de actividades y el asesoramiento al director.

El Instituto se encontraba dividido en secciones, órganos especializados para el estudio de la realidad española desde la óptica del régimen y en el ámbito de su competencia, así como para la elaboración de anteproyectos y dictámenes. Entre sus secciones, cuyo número varió

en el tiempo, podemos destacar la de Administración Pública, Leyes políticas, Filosofía social y sindical o Instituciones Europeas.

Había dos secretarías, una general que se ocupaba del régimen interior con departamentos como el de personal, archivo, registro o relaciones públicas e información. Y la secretaria técnica que se encargaba de la organización y proyección de las actividades y productos editoriales mediante departamentos como biblioteca, secretaría de revistas, intercambio cultural y ediciones. Existía una dirección de cursos y seminarios que se encargaba de la elaboración, organización y seguimiento del programa anual de cursos. Asimismo, se dispuso que los fondos de las bibliotecas del Congreso de los Diputados y del Senado pasaran a integrarse en su biblioteca. Finalmente, se contaba con una junta de coordinación integrada por el director, el secretario técnico, el secretario general y los jefes de departamento para el control de las actividades realizadas.

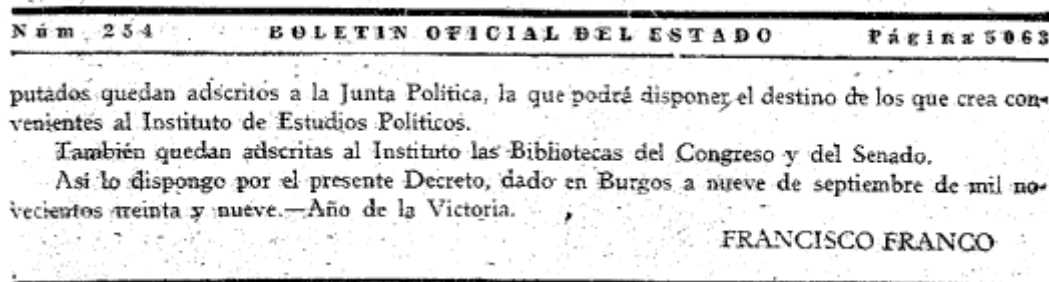


Figura 74. Disposición final por la que se ordena el depósito de los fondos de las Cortes en la biblioteca del nuevo instituto.

En 1977 se suprimió durante el periodo de la transición el instituto para unirse con el instituto de estudios administrativos IEA y pasar a llamarse CEPC. Trasladando su sede del Senado al palacio anexo de Godoy junto con la parte de fondos bibliográficos correspondiente a las adquisiciones efectuadas desde 1940⁷⁹⁶.

El instituto que no fue suprimido y que cambió su denominación por Centro de Estudios Constitucionales (CEC) fue el Instituto de Estudios Administrativos (IEA). El IEA no tenía el marcado carácter político del IEP lo cual hizo posible su pervivencia durante la transición al no estar tan ligado al viejo régimen. Sin embargo, a fin de unificar recursos y

⁷⁹⁶ Id.

evitar duplicidades se le incorporaron los recursos del IEP para constituir el CEC⁷⁹⁷. Así se constituyó por Decreto 2761/1977 de 28 de octubre el Centro de Estudios Constitucionales en el marco de un conjunto de medidas de reorganización de la presidencia del gobierno.

El IEA era una unidad especializada del centro de formación y perfeccionamiento de funcionarios, dependiente de la presidencia del gobierno, al que le confirió el carácter de organismo autónomo en la ley de jefatura del Estado nº93/61, de 23 de diciembre. Sus funciones principales eran la selección, formación y perfeccionamiento de los funcionarios. En 1966 el Centro cambió de nombre y pasó a llamarse Escuela Nacional de Administración Pública (EAP) de acuerdo con lo dispuesto en el artº 1 de su Reglamento Orgánico⁷⁹⁸, pasando el IEA a ser una de las unidades organizativas básicas de la EAP dedicada a promover estudios de investigación sobre distintos aspectos de la administración.

B.O. del E.—Núm. 191

11 agosto 1966

10401

REGLAMENTO ORGANICO DEL CENTRO DE FORMACION Y PERFECCIONAMIENTO DE FUNCIONARIOS

CAPITULO PRIMERO

FINES Y ACTIVIDADES

Artículo 1.º *Naturaleza, fines y sede de la institución.*

1. El Centro de Formación y Perfeccionamiento de Funcionarios, que podrá designarse en lo sucesivo Escuela Nacional de Administración Pública, es un Organismo autónomo creado por la Ley 93/1961, de 23 de diciembre, que depende de la Presidencia del Gobierno y tiene como fines la selección, formación y perfeccionamiento de los funcionarios públicos de acuerdo con lo dispuesto en el texto articulado de la Ley 109/1963, aprobado por Decreto 315/1964, de 7 de febrero.

2. El Centro, por determinación del Decreto 1140/1959, de 9 de julio, tiene su sede en el edificio de la antigua Universidad de Alcalá de Henarco, pudiendo, no obstante, cuando el cumplimiento de sus fines lo aconseje, desarrollar también sus tareas en otras localidades del país.

Art. 2.º *Actividades.*

Para el cumplimiento de sus fines compete al Centro organizar y desarrollar las actividades siguientes:

1. Pruebas selectivas en cualquiera de sus modalidades, y en todas las fases y trámites del correspondiente procedimiento, para ingreso en los Cuerpos Generales de la Administración

- a) Cursos generales de formación y perfeccionamiento.
- b) Cursos especiales y monográficos.
- c) Cursos abreviados y de divulgación.

Dentro de los Cursos especiales se dedicará particular atención a los que tengan por objeto materias de desarrollo económico.

2. De acuerdo con los distintos Ministerios, el Centro podría organizar cursos de formación y perfeccionamiento para funcionarios de Cuerpos Especiales.

3. Asimismo el Centro podrá concertar con otros organismos públicos la celebración de cursos específicos de formación o perfeccionamiento para el personal de éstos.

4. Sin perjuicio de admitir a titulados y funcionarios hispanoamericanos en cursos destinados a funcionarios españoles, el Centro organizará cursos exclusivos para aquéllos, en especial para personal directivo.

5. El Centro podrá organizar cursos destinados conjuntamente a funcionarios públicos y representantes de empresas y organizaciones privadas a los fines de lograr el conveniente intercambio de formación y de experiencias entre ambos sectores.

Art. 5.º *Estudios e investigaciones.*

El Centro promoverá o impulsará los estudios y trabajos de investigación sobre la Administración Pública mediante la concesión de ayudas económicas a funcionarios públicos y profesores o especialistas en las distintas materias, y a través de la labor que realicen en igual sentido los profesores y colaboradores del Centro.

Figura 75. Orden por la que se crea la Escuela de Administración Pública.

En 1977 traslada su sede al palacio de Godoy donde ya tenía su sede el IEP desde 1975. Se unieron así los recursos bibliográficos de ambos institutos si bien los fondos aportados por

⁷⁹⁷ DECRETO 2761/77, de 28 de octubre. Disposición final 4ª. En: <http://www.boe.es/boe/dias/1977/11/08/pdfs/A24430-24433.pdf>. Consultado el 15/02/2010.

⁷⁹⁸ ORDEN de 21 de julio de 1966. Reglamento orgánico. BOE nº191 de 11/08/66. En: <http://www.boe.es/datos/pdfs/BOE/1966/191/A10466-10469.pdf>. Consultado el 15/02/2010.

el IEA se integraron en el nuevo Instituto Nacional de Administración Pública (INAP) y sólo una pequeña parte pasó a integrar el fondo del CEPC, la nueva denominación desde 1977 del Centro de Estudios Constitucionales⁷⁹⁹.

El Centro de Estudios Políticos y Constitucionales (CEPC) tal como lo conocemos hoy en día nació en 1977 con el carácter de organismo autónomo adscrito a la Presidencia del Gobierno con los objetivos fundamentales de promover la investigación así como asesorar en materias constitucionales y sociales. No surge como organismo de nueva creación, sino que es el fruto de la necesidad de cambio y adaptación que durante la transición estaban viviendo todas las instituciones políticas y administrativas⁸⁰⁰.

Desde su creación en 1977 el CEPC ha conocido tres estructuras. La primera vino dada por una orden de presidencia del gobierno de 30 de enero de 1978 que fija la organización y funcionamiento del CEC como organismo dependiente directamente del ministerio de presidencia. Se le asignan ya entonces las funciones de realización y publicación de estudios constitucionales, prestación de asesoramiento al ministerio de presidencia en cuestiones constitucionales y administrativas y la elaboración de programas propios de investigación sobre sistemas constitucionales y administrativos. En general se hace hincapié en la investigación y asesoramiento con la finalidad de apoyar documentalmente a la recién constituida comisión constitucional encargada de una primera redacción de la que será constitución de 1978.

Una vez aprobada la constitución se normalizaron las actividades del CEPC de acuerdo con la estructura aprobada en la orden del ministerio de presidencia de 1977. Esta organización se mantuvo hasta 1980, año en el que se dictó el RD 1707/80 de 29 de agosto de organización y funcionamiento del CEC. En general se mantienen las características definidas en el decreto de 1977 aunque se realiza una importante reorganización estructural. Se mantienen las mismas funciones de la orden de 1978 y se le añaden las funciones de: i) formar y custodiar un fondo bibliográfico sobre Ciencia Política, Teoría del Estado y

⁷⁹⁹ ROSA IGLESIAS, B. El “Centro de Estudios Políticos y Constitucionales” de Madrid. *Historia Constitucional*, nº3, 2002. pp.312-314. En: <http://www.historiaconstitucional.com/index.php/historiaconstitucional/article/view/187/165>. Consultado el 15/02/2010.

⁸⁰⁰ Id.

Derecho Público, ii) Desarrollar ciclos y cursos especializados⁸⁰¹. Funciones que ya venía realizando los departamentos de Documentación y Estudios pero que al hacer expresa su mención refuerza el carácter del CEC como centro de formación e investigación.

PRESIDENCIA DEL GOBIERNO

19154

REAL DECRETO 1707/1980, de 29 de agosto, sobre organización y funcionamiento del Centro de Estudios Constitucionales.

Modificado el Real Decreto dos mil setecientos sesenta y uno/ mil novecientos sesenta y siete, de veintiocho de octubre, por el Real Decreto mil cuarenta y nueve/ mil novecientos ochenta, de seis de junio, que adscribe el Centro de Estudios Constitucionales al Ministro adjunto al Presidente, Encargado de la Coordinación Legislativa, y a la vista de la experiencia habida en el funcionamiento de dicho Centro desde la vigencia de la Orden ministerial de la Presidencia de treinta de enero de mil novecientos setenta y ocho, procede dotar al Centro de Estudios Constitucionales de un nuevo régimen orgánico y funcional.

En su virtud, a propuesta del Ministro de la Presidencia, con informe del Ministerio de Hacienda y previa deliberación del Consejo de Ministros en su reunión del día veintinueve de agosto de mil novecientos ochenta,

DISPONGO:

Artículo primero.—El Centro de Estudios Constitucionales, Organismo Autónomo adscrito al Ministro adjunto al Presidente, Encargado de la Coordinación Legislativa, tiene encomendadas las siguientes funciones:

- Elaborar y promover tareas de investigación sobre los caracteres, evolución y desarrollo de las instituciones de los sistemas sociales, políticos, constitucionales y administrativos, en su proyección nacional e internacional.

- Realizar, promover y publicar estudios sobre cualesquiera materias relacionadas con las áreas de investigación a que se refiere el precedente apartado.
- Formar y custodiar un fondo documental y bibliográfico sobre Ciencia Política, Teoría del Estado, Derecho Público y materias conexas.
- Desarrollar ciclos y cursos de enseñanzas especializadas.
- Prestar asistencia y asesoramiento a la Presidencia del Gobierno.
- Realizar cuantos cometidos se le encomienden por el Ministro a quien el Centro está adscrito.

Artículo segundo.—Son órganos de gobierno y administración del Centro de Estudios Constitucionales:

- El Consejo Rector.
- El Director.
- Los Subdirectores generales de Estudios e Investigación y de Publicaciones y Documentación.
- El Gerente, con categoría de Subdirector general.

Artículo tercero.—Uno. Compete al Consejo Rector elaborar las directrices de actuación del Centro y velar por su cumplimiento.

Dos. El Consejo Rector del Centro de Estudios Constitucionales tendrá la composición siguiente:

Presidente: El Ministro adjunto al Presidente, Encargado de la Coordinación Legislativa.

Vicepresidente: El Director del Centro de Estudios Constitucionales.

Vocales:

- El Presidente de la Real Academia de Ciencias Morales y Políticas.
- El Presidente de la Comisión General de Codificación.

Figura 76. Real Decreto sobre organización y funcionamiento del Centro de Estudios Constitucionales.

Entre los cambios habidos en el organigrama del CEC nos parece interesante destacar la desaparición del Departamento de Relaciones e Intercambio pasando a asumir sus funciones el nuevo Departamento de Estudios y Documentación. Además se crea la Subdirección de Estudios e Investigación con las funciones de programar y realizar proyectos de investigación así como la realización de actividades docentes y culturales. Se crea también la Subdirección General de Publicaciones y Documentación con las funciones de edición y distribución de publicaciones, la conservación de los fondos bibliográficos y la explotación y difusión de datos relacionados con los resultados de los proyectos de investigación. Del Departamento de Documentación dependen la Sección de Documentación Científica, la de Fondos Documentales y la de Documentación Informativa.

⁸⁰¹ REAL DECRETO 1707/80, de 29 de agosto de 1980. Artº1º. En: <http://www.boe.es/boe/dias/1980/09/04/pdfs/A19942-19943.pdf>. Consultado el 15/02/2010.

En 1997 se produce la última reorganización del Centro, regulado por el Real Decreto 1269/97 por el cual se mantiene la misma estructura pero se modifica su denominación que pasa de Centro de Estudios Constitucionales a Centro de Estudios Políticos y Constitucionales, más comprensiva de sus competencias que se redefinen en la norma, pretendiendo acomodar su estructura a la nueva normativa administrativa de la Administración General del Estado al tiempo que potenciar sus funciones en el marco de la Unión Europea y los países iberoamericanos⁸⁰².

También se establecen algunos cambios en la composición de los órganos rectores y se añade la función de⁸⁰³:

conceder especial atención, en el desarrollo de sus funciones, a las cuestiones relativas a las instituciones propias de los países iberoamericanos y a las relaciones de éstos con España y Europa.

2. Los equipos de protección individual utilizados como prendas de vestir o por encima de dichas prendas deberán ser de colores vivos, contrastar con el medio marino y ser bien visibles.

17826 REAL DECRETO 1269/1997, de 24 de julio, por el que se regula la organización y funcionamiento del Centro de Estudios Políticos y Constitucionales.

Transcurridos más de dieciséis años desde la entrada en vigor del Real Decreto 1707/1980, de 29 de agosto, por el que se regulaba la organización y funcionamiento del Centro de Estudios Constitucionales, y teniendo en cuenta la experiencia adquirida desde entonces, se estima oportuno promulgar una nueva norma en la que se redefinan sus funciones y competencias, y, de manera particular, resaltar la especial atención que merecen los temas relacionados con las instituciones propias de los países iberoamericanos, y a las relaciones de éstos con España en sus materias objeto de estudio e investigación.

A tal efecto, se modifica su denominación, que pasaría a ser la de Centro de Estudios Políticos y Constitucionales, como más comprensiva de sus competencias, para lo que resulta oportuno mejorar la eficacia del organismo y potenciar las funciones de sus órganos rectores y de administración.

Por otra parte, debe tenerse en cuenta que la aprobación de la Ley 6/1997, de 14 de abril, de Organización y Funcionamiento de la Administración General del Estado, ha supuesto un nuevo régimen regulador de los Organismos autónomos, determinándose en su disposición transitoria tercera la necesidad de adaptar los organismos existentes a la regulación de la citada Ley.

El presente Real Decreto da cumplimiento a dicha previsión y procede a la adaptación del Centro de Estudios Políticos y Constitucionales al tipo de Organismo autónomo, regulado en la Ley 6/1997, la cual ha de llevarse a cabo por Real Decreto, a propuesta conjunta de los Ministros de Administraciones Públicas y de Economía y Hacienda y de acuerdo con el Ministerio del que depende el organismo, ya que, en este caso, la adaptación no

DISPONGO:

Artículo 1. Naturaleza y régimen jurídico.

1. El Centro de Estudios Políticos y Constitucionales es un Organismo autónomo de los previstos en el artículo 43.1.a) de la Ley 6/1997, de 14 de abril, de Organización y Funcionamiento de la Administración General del Estado, con personalidad jurídica y patrimonio propios, adscrito al Ministerio de la Presidencia, cuya finalidad es el estudio y la investigación de los sistemas sociales, políticos, constitucionales y administrativos, manteniendo a tales efectos las oportunas relaciones con las instituciones propias de los países iberoamericanos.

El Ministerio de la Presidencia ejercerá respecto del Centro el control de eficacia en los términos previstos en el artículo 51 de la Ley 6/1997.

2. El Centro de Estudios Políticos y Constitucionales se rige por la Ley 6/1997, de 14 de abril, de Organización y Funcionamiento de la Administración General del Estado; por la Ley General Presupuestaria, texto refundido aprobado por Real Decreto legislativo 1091/1988, de 23 de septiembre; por el presente Real Decreto, y demás disposiciones de aplicación a los Organismos autónomos de la Administración General del Estado.

Artículo 2. Funciones.

Para el cumplimiento de sus fines, el Centro tiene encomendadas las funciones que a continuación se relacionan, sin perjuicio de las funciones de formación e investigación que en estas materias estén atribuidas a otros organismos de la Administración General del Estado:

1. Elaborar y promover tareas de estudio e investigación sobre el carácter, evolución y funcionamiento de los sistemas sociales, políticos, constitucionales y administrativos, tanto en su dimensión nacional como internacional.

2. Realizar, promover y publicar, en su caso, estudios sobre cualesquiera materias relacionadas con las áreas a que se refiere el precedente apartado.

Figura 77. Real Decreto sobre organización y funcionamiento del Centro de Estudios Políticos y Constitucionales.

⁸⁰² Rosa Iglesias, B. Op. cit., p.314.

⁸⁰³ Artículo nº2, apartado 6 del Real Decreto 1269/97, de 7 de agosto de 1997.

En: <http://www.boe.es/boe/dias/1997/08/07/pdfs/A24078-24080.pdf>. Consultado el 15/02/2010.

En el artículo 1º del Real Decreto 1269/1997 de 24 de julio se dispone que el Centro es un organismo autónomo con personalidad jurídica y patrimonio propios adscrito al Ministerio de Presidencia. En su artículo 2º es donde se especifican sus funciones, que recogemos literalmente desde el documento que en el sitio web del CEPC recoge este Real Decreto⁸⁰⁴:

- *Elaborar y promover tareas de estudio e investigación sobre el carácter, evolución y funcionamiento de los sistemas sociales, políticos, constitucionales y administrativos tanto en su dimensión nacional como internacional.*
- *Realizar, promover y publicar, en su caso, estudios sobre cualesquiera materias relacionadas con las áreas a que se refiere el precedente apartado.*
- *Formar y custodiar un fondo documental y bibliográfico sobre Derecho Constitucional, Teoría del Estado, Teoría de la Constitución, Ciencia Política, Historia de las Ideas Políticas e Historia Política de España, y materias conexas de Derecho Público.*
- *Desarrollar ciclos y cursos de enseñanzas especializadas en las materias citadas en los apartados precedentes.*
- *Prestar asistencia y asesoramiento a la Presidencia del Gobierno.*
- *Conceder especial atención, en el desarrollo de sus funciones, a las cuestiones relativas a las instituciones propias de los países iberoamericanos y a las relaciones de éstos con España y Europa.*
- *Realizar cuantos cometidos se le encomienden por el Ministerio al que el Centro está adscrito.*

La estructura organizativa básica del CEPC gira en torno a las Subdirecciones Generales, de entre las cuales es de nuestro interés la correspondiente al área de Publicaciones y

⁸⁰⁴ Artículo nº2 del Real Decreto 1269/97, de 7 de agosto de 1997.

En: <http://www.boe.es/boe/dias/1997/08/07/pdfs/A24078-24080.pdf>. Consultado el 15/02/2010.

Documentación, a quien corresponde formular la propuesta de programa editorial y la dirección y organización de Biblioteca y Archivo⁸⁰⁵.

6.2.2. La biblioteca y Centro de Documentación del CEPC a través de su sitio web.

El CEPC dispone de una de las bibliotecas más completas en su especialidad, el Derecho Público y la Ciencia Política⁸⁰⁶. En el fondo se encuentran más de 70.000 títulos de monografías sobre Derecho Constitucional, Ciencia Política, Historia Política de España y Derecho Público. Además cuenta con más de 1500 títulos de revistas españolas y extranjeras de las que 730 tienen viva su suscripción⁸⁰⁷.

Asimismo, cuenta con un importante servicio editorial que publica diez colecciones de monografías y seis títulos de revistas tanto en soporte digital como impreso entre las que se cuentan revistas tan conocidas como la “Revistas de Estudios Políticos” o la “Revista Española de Derecho Constitucional”⁸⁰⁸.

Existen dos vías de acceso a las publicaciones del CEPC. I) Un modo es a través de la pestaña “Publicaciones”, como se muestra en la figura de a continuación. Esta opción permite el acceso a los libros y revistas editados por el propio centro, así como proporcionar una pequeña biblioteca digital de los artículos publicados en las revistas del Centro⁸⁰⁹. En el margen izquierdo se muestra el índice de todos los materiales a los que se accede desde esta pestaña. Entre ellos se encuentra el listado de revistas con el acceso electrónico, o información sobre colecciones.

⁸⁰⁵ Ibid., art. nº6.

⁸⁰⁶ ROSA IGLESIAS, B. Op. cit., p.316.

⁸⁰⁷ MEMORIA 2008. *Centro de Estudios Políticos y Constitucionales*. p.55.

En: http://www.cepc.es/include_mav/getfile.asp?IdFileImage=3241. Consultado el 15/02/2010.

⁸⁰⁸ Todas ellas evaluadas en la tesis doctoral de GÓMEZ DEL PULGAR RODRÍGUEZ DE SEGOVIA, G. *Calidad de las Revistas Españolas de Derecho Público*. [Tesis Doctoral]. Director: JOSÉ RAMÓN PÉREZ ÁLVAREZ OSSORIO. Tutor: ELÍAS SANZ CASADO. Departamento de Biblioteconomía y Documentación. Universidad Carlos III. En: <http://e-archivo.uc3m.es/bitstream/10016/495/10/Gomez%20del%20Pulgar%20Rodriguez%20de%20Segovia.%20Gloria%281%29.pdf>.

Consultado el 15/02/2010.

⁸⁰⁹ *Publicaciones del CEPC*. En: <http://www.cepc.es/colecciones.asp>. Consultado el 15/02/2010.

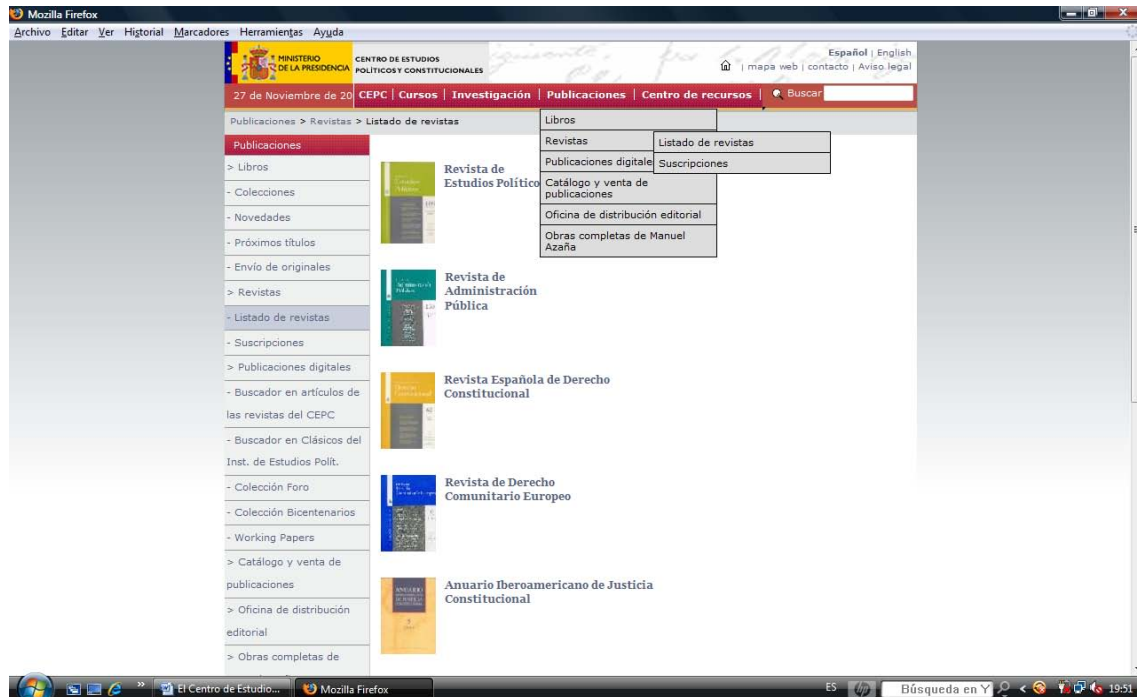


Figura 78. Página web del listado de revistas publicadas por el CEPC en la pestaña “Publicaciones”.

Asimismo, cuenta con un enlace a un motor de búsqueda denominado RAP que permite la búsqueda retrospectiva de artículos publicados en las revistas del Centro por campos como “autor”, “título del artículo” o “sección”, entre otros. También se incluyen formularios de búsqueda por los números de cada revista, sus secciones y periodos de publicación. Si bien no permite la búsqueda por materias, si es posible hacerlo por palabras en texto libre. Opción que se ve considerablemente mejorada si los usuarios emplean el tesoro del Ministerio de Presidencia con que cuentan. A continuación mostramos un ejemplo de búsqueda en la Revista Española de Derecho Constitucional de la crítica de un libro sobre Kelsen escrito por un autor con apellido “Barba” como mostramos en la figura siguiente.

RAP - Mozilla Firefox

http://www.cepc.e...nicio_libros.asp# RAP RAP

MINISTERIO DE LA PRESIDENCIA CENTRO DE ESTUDIOS POLÍTICOS Y CONSTITUCIONALES

Revista: Revista Española de Derecho Constitucional (1981 - 2000)
(Números disponibles: del 1 al 60)

Título del artículo: kelsen

Autor: barba

Sección: CRITICA DE LIBROS

Número: Desde: Hasta:

Año: Desde: Hasta:

Palabras en el texto: justicia

Contiene: Todas las palabras

buscar nueva consulta

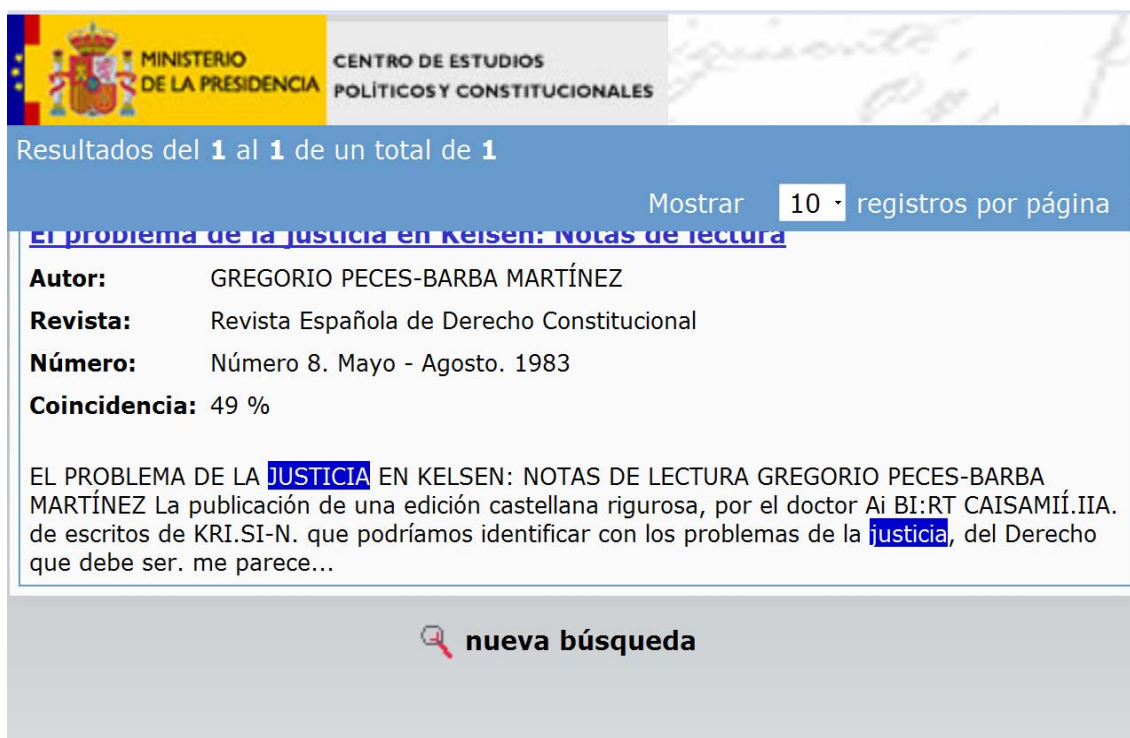
Acceso directo al número: 8 abrir

ayuda

El Centro de Estudio... RAP - Mozilla Firefox ES Búsqueda en Y 20:44

Figura 79. Aplicación RAP empleada para la búsqueda de un artículo.

El resultado de la búsqueda es una referencia como la que sigue a una reseña del profesor Peces-Barba titulado: “El problema de la justicia en Kelsen: Notas de lectura”. Además de indicar la referencia bibliográfica de la obra, se indica el grado de coincidencia del resultado de la búsqueda respecto a la estrategia de búsqueda empleada. Como vemos la coincidencia es tan sólo del 49%, lo cual nos indica que debimos refinar más la búsqueda. Asimismo, se incluye el comienzo de la reseña, pero normalmente para los artículos se incluye su resumen.



MINISTERIO DE LA PRESIDENCIA CENTRO DE ESTUDIOS POLÍTICOS Y CONSTITUCIONALES

Resultados del 1 al 1 de un total de 1

Mostrar 10 registros por página

El problema de la justicia en Kelsen: Notas de lectura

Autor: GREGORIO PECES-BARBA MARTÍNEZ

Revista: Revista Española de Derecho Constitucional

Número: Número 8. Mayo - Agosto. 1983

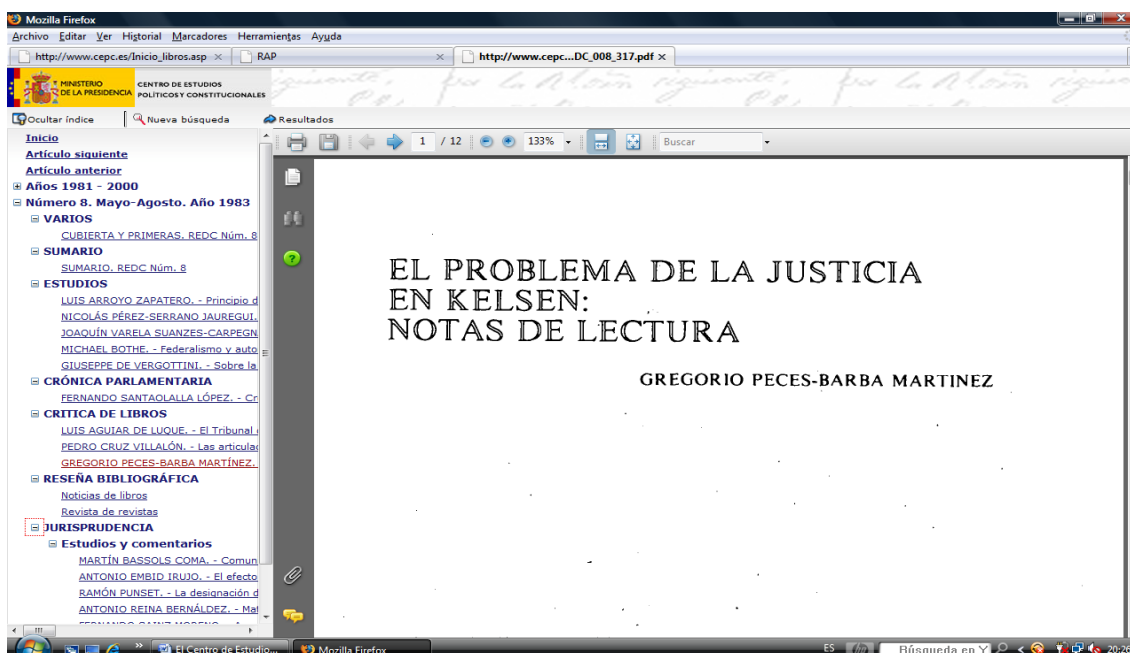
Coincidencia: 49 %

EL PROBLEMA DE LA JUSTICIA EN KELSEN: NOTAS DE LECTURA GREGORIO PECES-BARBA MARTÍNEZ La publicación de una edición castellana rigurosa, por el doctor AIBI:RT CAISAMIÍ.IIA. de escritos de KRI.SI-N. que podríamos identificar con los problemas de la justicia, del Derecho que debe ser. me parece...

nueva búsqueda

Figura 80. Resultado de la búsqueda del artículo.

El título es un hipervínculo que permite al usuario recuperar el texto completo del trabajo junto con un índice en su margen izquierdo del sumario de la revista donde se publicó.



Mozilla Firefox

http://www.cepc.es/Inicio_libros.asp RAP http://www.cepc.es/DC_008_317.pdf

MINISTERIO DE LA PRESIDENCIA CENTRO DE ESTUDIOS POLÍTICOS Y CONSTITUCIONALES

Nueva búsqueda Resultados

Inicio
Artículo siguiente
Artículo anterior
Años 1981 - 2000
Número 8. Mayo-Agosto. Año 1983
VARIOS
CUBIERTA Y PRIMERAS. REDC Núm. 8
SUMARIO. REDC Núm. 8
ESTUDIOS
LUIS ARROYO ZAPATERO. - Principio d
NICOLÁS PÉREZ-SERRANO JAUREGUI
JOAQUÍN VARELA SUANZES-CARPEÑ
MICHAEL ROTH. - Federalismo y auto
GIUSEPPE DE VERGOTTINI. - Sobre la
CRÓNICA PARLAMENTARIA
FERNANDO SANTIAGALLA LÓPEZ. - Cr
CRÍTICA DE LIBROS
LUIS AGUIAR DE LUQUE. - El Tribunal
PEDRO CRUZ VILLALÓN. - Las artícu
GREGORIO PECES-BARBA MARTÍNEZ
RESEÑA BIBLIOGRÁFICA
Noticias de libros
Revista de revistas
JURISPRUDENCIA
Estudios y comentarios
MARTÍN BASSOLS COMA. - Comun
ANTONIO EMBID IRUJO. - El efecto
RAMÓN PUNSET. - La designación d
ANTONIO REINA BERNÁLDEZ. - Ma
FERNANDO SANTIAGALLA LÓPEZ. - Cr

EL PROBLEMA DE LA JUSTICIA
EN KELSEN:
NOTAS DE LECTURA

GREGORIO PECES-BARBA MARTINEZ

Figura 81. Recuperación del artículo a texto completo.

II) El segundo modo de acceso es a través de la pestaña “Centro de Recursos” con el cual el usuario accede a un importante conjunto de publicaciones oficiales españolas y

extranjeras de contenido parlamentario, legislativo y jurisprudencial⁸¹⁰. Asimismo, ofrece referencias de un importante fondo documental de la Unión Europea y del Consejo de Europa, así como varias bases de datos bibliográficas y jurídicas. Los fondos que custodian ambos servicios del Área de Biblioteca y Documentación Jurídica son de libre acceso tanto físicamente como a través del sitio web.

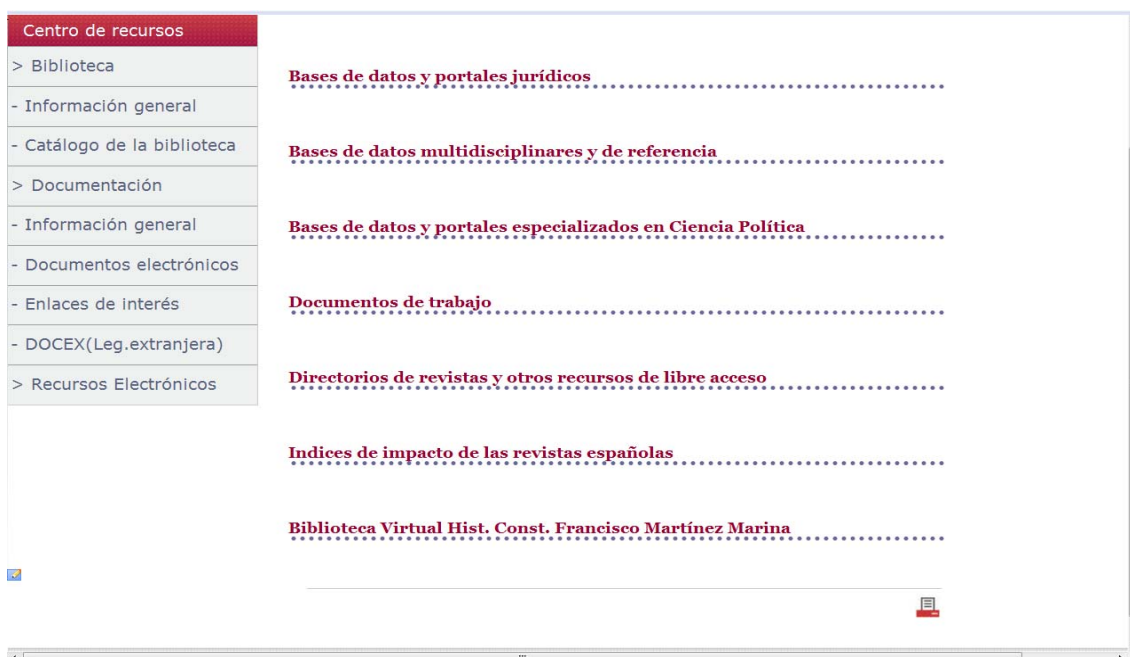


Figura 82. Centro de Recursos del CEPC.

Respecto al sitio web, debemos comentar que se trata de un sitio bien estructurado, con un nivel AA en criterios de accesibilidad de la Web Accessibility Initiative. Está orientado no sólo a la difusión de las actividades del Centro, sino que recoge numerosos servicios de valor añadido tales como un catálogo, bases de datos o incluso acceso a Bibliotecas digitales como la biblioteca de Historia Constitucional Francisco Martínez Marina. En 1997, aprovechando la última reforma del centro, se puso en marcha una página web institucional a través de la cual cualquier usuario podía acceder al catálogo del Centro sin necesidad de encontrarse en el mismo ni estar registrado como tal⁸¹¹. Además, se permite el acceso a un catálogo de publicaciones periódicas, a los dossiers publicados en su Boletín de Documentación, a las actividades del organismo y a toda aquella información relacionada

⁸¹⁰ Centro de Recursos del CEPC. En: http://www.cepc.es/recursos_electronicos.asp. Consultado el 15/02/2010.

⁸¹¹ ROSA IGLESIAS, B. Op. cit., p.317.

con las actividades y cursos ofrecidos por el centro de acuerdo con la carta de servicios electrónicos ofrecidos por el centro y publicados en el propio sitio web del CEPC⁸¹².

En cuanto al aspecto docente, los cursos⁸¹³ ofrecidos por el CEPC son coordinados por la Subdirección general de Estudios e Investigación, a quien corresponde programar, promover y realizar estudios, informes y proyectos de investigación, así como llevar a cabo las actividades docentes del Centro, especialmente del Curso de Derecho Constitucional y Ciencia Política. La finalidad de este curso es contribuir a la especialización de postgraduados en las áreas de Derecho Constitucional y Ciencia Política. El curso, cuya primera edición data de 1979, se ha convertido en un Diploma de Derecho Constitucional y Ciencia Política que es reconocido por los estudios de postgrado de numerosas universidades entre las que se encuentra la Universidad Carlos III⁸¹⁴.

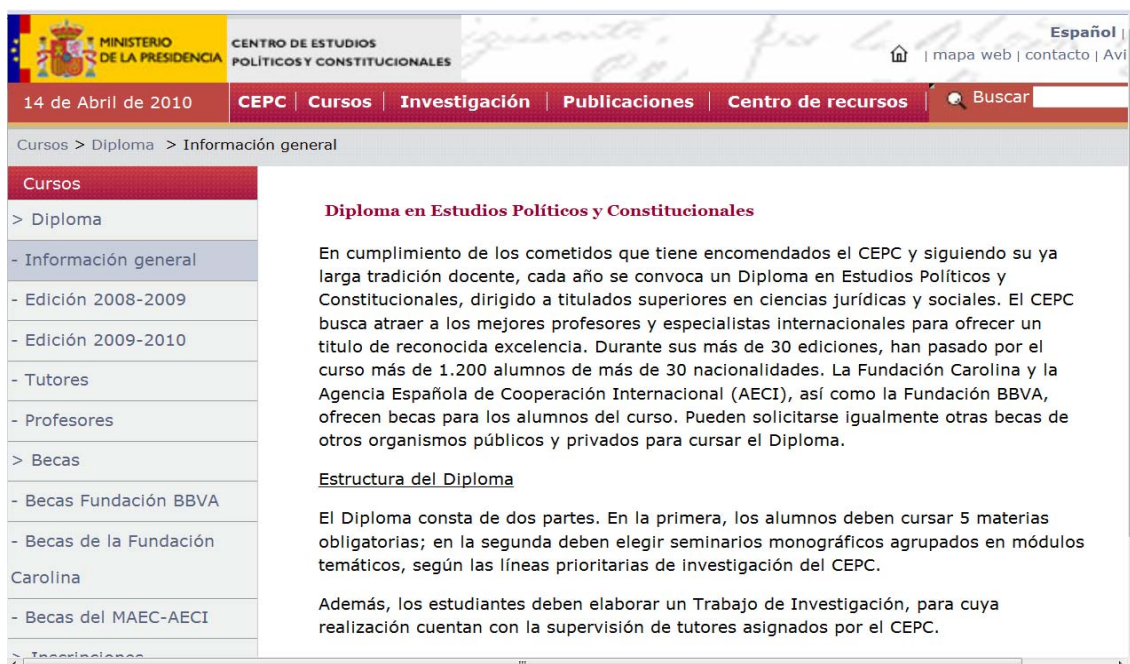


Figura 83. Página web con información sobre el Diploma de Estudios Políticos y Constitucionales con su correspondiente mapa de sitio en el margen izquierdo de la página.

Además de este curso, el Centro organiza Conferencias y Seminarios monográficos a lo largo del año impartidos por especialistas en la materia, que se anuncian en el propio sitio web del CEPC o bien se puede solicitar un servicio de alerta informativa que por medio de

⁸¹² CARTA DE SERVICIOS ELECTRÓNICOS. En: http://www.cepc.es/include_mav/getfile.asp?IdFileImage=280. Consultado el 15/02/2010.

⁸¹³ Cursos del CEPC. En: http://www.cepc.es/informacion_diploma.asp. Consultado el 15/02/2010.

⁸¹⁴ Ibid., p.318.

correo electrónico permite tener a los usuarios informados acerca de la realización de estos eventos.

6.3. Evaluación global de la biblioteca digital por parte de los usuarios del CEPC.

A lo largo de una semana hemos acudido a la biblioteca del CEPC para solicitar a sus usuarios la realización de una encuesta electrónica diseñada con el programa Google Docs. Se les ofreció la posibilidad de realizarla en el mismo centro o bien se les pedía el correo electrónico a donde se les enviaba.

Se ha diseñado un cuestionario cuyo ámbito ha sido el CEPC. El universo de la muestra son los usuarios del centro de ambos sexos y mayores de 18 años. La muestra tiene un tamaño de 30 entrevistas de las que llegamos a realizar 27. No procede ponderación en la recogida. El punto de muestreo ha sido la biblioteca del centro y sus aulas. El procedimiento de muestreo ha sido polietápico con selección de los individuos por rutas aleatorias y cuotas de alumnos, profesores y personal del centro. Los cuestionarios se han aplicado mediante entrevista personal en la biblioteca y las aulas. La fecha de realización fue entre el 04 de mayo y el 11 de mayo de 2008.

Respecto al error muestral⁸¹⁵, al tratarse de un muestreo aleatorio simple, y teniendo en cuenta que la variable a estimar es una proporción, entonces el error de la muestra lo calculamos de acuerdo con la ecuación $e=2 * \sqrt{p} * q \div n$. Como el error muestral que vamos a calcular lo queremos con un nivel de confianza del 95%, hemos multiplicado el algoritmo anterior por 2. Para nuestra encuesta vamos a considerar que $p=q$ (porcentajes de que sean o no una característica) e igual a 50% porque por igual es la probabilidad del resultado p y q . “ n ” es la muestra que para nuestro caso fueron únicamente 30 usuarios. Aceptando todas estas condiciones, diremos que el error muestral de la encuesta es el siguiente para un error muestral del 95%.

$$e=2 * \sqrt{50} * 50 \div 30 = \pm 18\%$$

⁸¹⁵ ALVIRA MARTÍN, F. *La encuesta: una perspectiva general metodológica*. Madrid : Centro de Investigaciones Sociológicas, 2004. pp.87-88.

El cuestionario consta de nueve preguntas, las cinco primeras acerca del perfil de usuario así como la frecuencia con que acude a la biblioteca del CEPC. Las siguientes cuatro preguntas tienen que ver con el grado de conocimiento que tienen de los recursos digitales que el centro pone a disposición de los usuarios a través de su sitio web. Para el diseño de la encuesta nos hemos basado en el trabajo de Azofra sobre técnicas de diseño de cuestionarios⁸¹⁶, así como en la tesis doctoral de Kruk sobre evaluación de bibliotecas digitales semánticas⁸¹⁷. A continuación mostramos la tabla con las preguntas realizadas en el cuestionario.

¿En qué intervalo de edad se encuentra usted?	¿Cuál es su nivel más alto de estudios?	¿En qué lenguas lee usted?	¿Cuál es su ámbito de formación?	¿Con qué frecuencia acude a la biblioteca del Centro de Estudios Políticos y Constitucionales?	¿Con qué frecuencia utiliza los recursos digitales de la biblioteca desde el sitio web del centro?	¿Cuál es su impresión general acerca del sitio web del CEPC?	¿Cómo le ha resultado la tarea de localizar artículos de derecho constitucional sobre el Senado?	Localice el año de publicación de una obra de Álvarez Junco sobre Lerroux y responda cómo le ha parecido la búsqueda.
Entre 35 y 65 años	Postgrado	Español, Inglés, Francés, Portugués	Historia Contemporánea	Semanalmente	Semanalmente	Difícil de utilizar, Bien organizado, Poco intuitivo, Adecuado a las necesidades	Difícil de buscar, Obtuve muchos resultados, Obtuve resultados poco pertinentes	Difícil de buscar
Entre 35 y 65 años	Postgrado	Español, Inglés, Francés, Portugués	Estudios Africanos, Sociología	Diariamente	Diariamente	Fácil de utilizar, Bien organizado, Intuitivo, Adecuado a las necesidades	Fácil de buscar, Obtuve muchos resultados, Obtuve resultados precisos	Fácil de buscar, Obtuve resultados precisos
Mayor de 65 años	Doctor	Español, Francés, Italiano	Derecho Civil, Derecho Constitucional	Diariamente	Diariamente	Difícil de utilizar, Mal estructurado, Poco intuitivo, Adecuado a las necesidades	No he sido capaz de encontrar nada	No he sido capaz de encontrar nada
Entre 18 y 23 años	Bachiller	Español, Inglés	Ciencias Políticas	Mensualmente	Mensualmente	Difícil de utilizar, Bien organizado, Poco intuitivo, Adecuado a las necesidades	No he sido capaz de encontrar nada	Difícil de buscar, Obtuve muchos resultados, Obtuve resultados poco pertinentes
Entre 23 y 35 años	Licenciado	Español, Inglés	Historia del Derecho, Derecho Civil, Derecho Constitucional, Derecho Internacional, Derecho Comunitario	Semanalmente	Semanalmente	Difícil de utilizar, Bien organizado, Poco intuitivo, Adecuado a las necesidades	Difícil de buscar, Obtuve muchos resultados, Obtuve resultados poco pertinentes	No he sido capaz de encontrar nada
Entre 23 y 35 años	Licenciado	Español, Francés	Historia del Derecho, Derecho Civil, Derecho Constitucional, Estudios Iberoamericanos, Historia Contemporánea	Semanalmente	Semanalmente	Difícil de utilizar, Bien organizado, Poco intuitivo, Adecuado a las necesidades	Difícil de buscar, Obtuve muchos resultados, Obtuve resultados poco pertinentes	Fácil de buscar, Obtuve resultados precisos
Entre 35 y 65 años	Licenciado	Español, Inglés, Francés	Ciencias Políticas, Estudios Iberoamericanos	Semanalmente	Semanalmente	Difícil de utilizar, Bien organizado, Poco intuitivo, Adecuado a las necesidades	No he sido capaz de encontrar nada	Difícil de buscar, Obtuve pocos resultados, Obtuve resultados poco pertinentes

Figura 84. Encuesta digital realizada entre los usuarios. Elaboración propia.

6.3.1. Análisis de resultados de la encuesta.

1. En la primera pregunta, sobre el intervalo de edad de nuestros usuarios, se muestra que existe una clara mayoría de personas entre los 23 y 35 años por los investigadores y

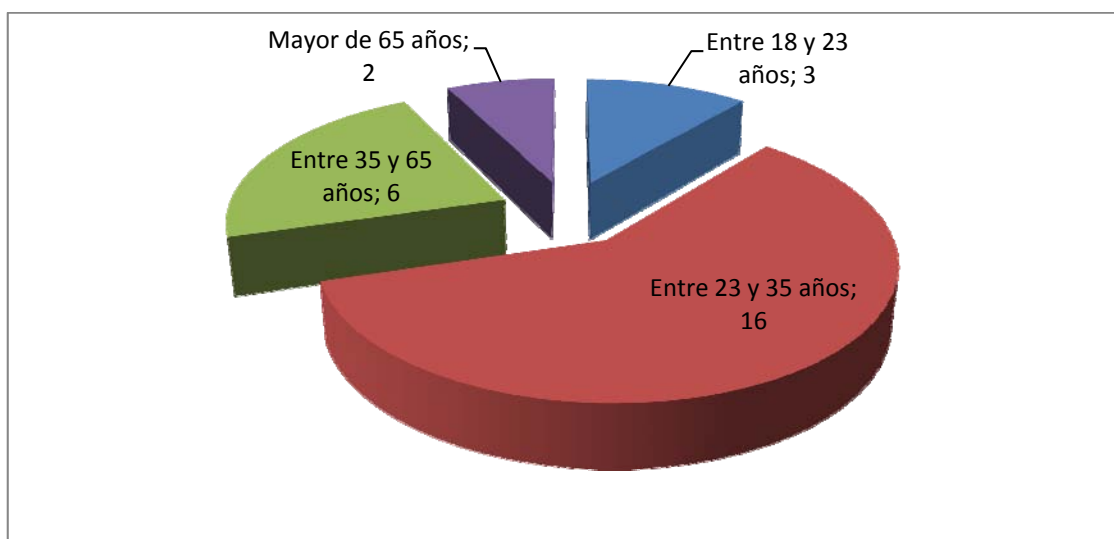
⁸¹⁶ AZOFRA MÁRQUEZ, M^a J. *Cuestionarios*. Madrid : CIS, 1999.

⁸¹⁷ RYSZARD KRUK, S. *Semantic Digital Libraries. Improving usability of information discovery with semantic and social services*. Dublin, 2010.

alumnos del “Diploma de Estudios Políticos y Constitucionales”. El segundo grupo en importancia ha sido el correspondiente al grupo de edad entre los 35 y 65 años que comprende fundamentalmente a profesores del diploma y personal de administración del centro. Asimismo, existe un grupo de estudiantes de las universidades con las que el CEPC tiene suscritos acuerdos para que sus alumnos puedan acudir al mismo. Finalmente acuden un pequeño número de investigadores mayores de 65 años que son profesores o bien profesionales del ámbito jurídico jubilados.

Entre 18 y 23 años	3	11%
Entre 23 y 35 años	16	59%
Entre 35 y 65 años	6	22%
Mayor de 65 años	2	7%

Tabla 11. Tramo de edad de los usuarios encuestados. Elaboración propia.



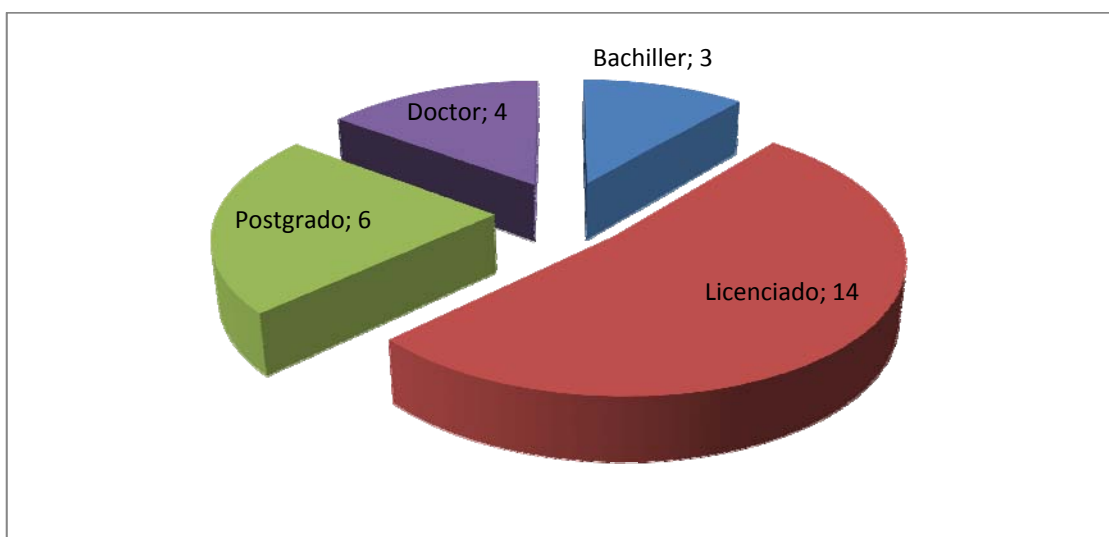
Gráfica 1. Gráfica correspondiente a los tramos de edad de la comunidad de usuarios. Elaboración propia.

2. Sobre la segunda pregunta, referida a la posesión del título educativo de nivel más elevado, nos encontramos con que una gran mayoría de los mismos están en posesión de al menos un título de licenciado. Hecho lógico si pensamos que la mayoría de nuestros usuarios son alumnos del postgrado del CEPC. El segundo grupo por orden de importancia corresponde a alumnos con algún título de postgrado, lo que se produce fundamentalmente entre alumnos provenientes de países iberoamericanos que han venido a estudiar el diploma del CEPC. Un tercer grupo corresponde a los profesores del Centro e investigadores que acuden desde otras universidades a utilizar el importante fondo

hemerográfico del CEPC. Finalmente, contamos con los alumnos que están estudiando una carrera y que acuden para la realización de algún trabajo, fundamentalmente se trata de alumnos de derecho que desean consultar las revistas editadas por el centro. Al no ser todavía licenciados los hemos considerado como usuarios con el grado de bachiller.

Bachiller	3	11%
Licenciado	14	52%
Postgrado	6	22%
Doctor	4	15%

Tabla 12. Nivel de estudios de los usuarios encuestados. Elaboración propia.



Gráfica 2. Gráfica correspondiente a los niveles educativos de la comunidad de usuarios. Elaboración propia.

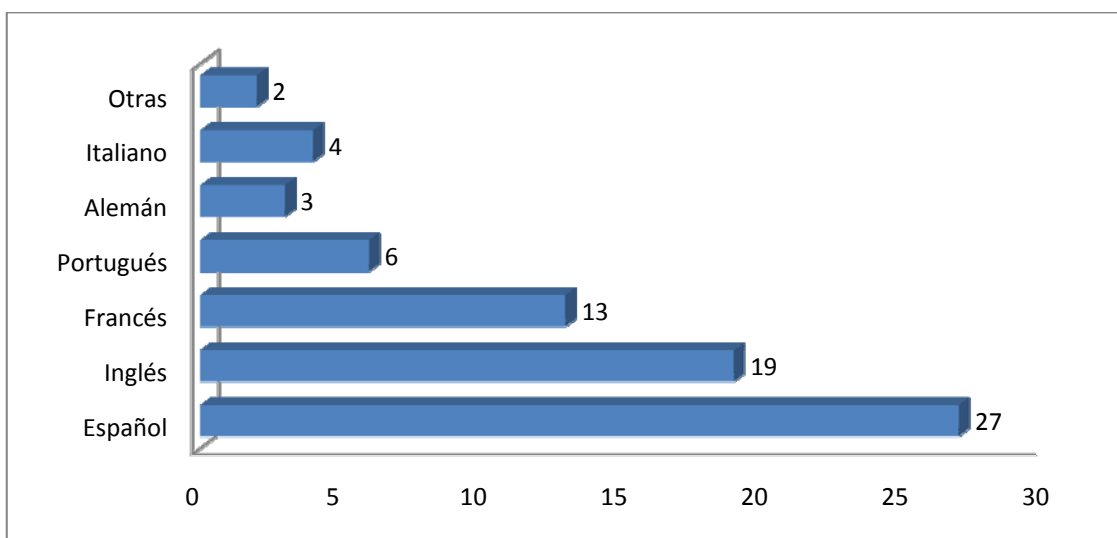
3. Respecto a las lenguas de comprensión lectora, hemos incluido esta pregunta porque la colección incluye textos en otros idiomas. Esto nos permitirá en primer lugar seleccionar las lenguas en las que consultar los documentos digitales de la colección y en segundo lugar asistir a los bibliotecarios en la toma de decisiones acerca de qué documentos de la colección digitalizar para la biblioteca digital semántica.

De acuerdo con los datos de la encuesta todos nuestros usuarios comprenden textos en español y una gran mayoría dicen ser capaces de comprender textos en inglés. Sin embargo, llama la atención el escaso número de usuarios capaces de leer en francés, italiano o alemán. Esto supone un serio problema porque la tradición jurídica española tiene poco que ver

con la anglosajona y si mucho con la francesa en áreas como el derecho civil o la alemana en áreas como el derecho constitucional. Quizás para las áreas de historia contemporánea o ciencia política pueda ser más pertinente la digitalización de textos en inglés.

Español	27	100%
Inglés	19	70%
Francés	13	48%
Portugués	6	22%
Alemán	3	11%
Italiano	4	15%
Otras	2	7%

Tabla 13. Lenguas de comprensión lectora para de los usuarios encuestados. Elaboración propia.



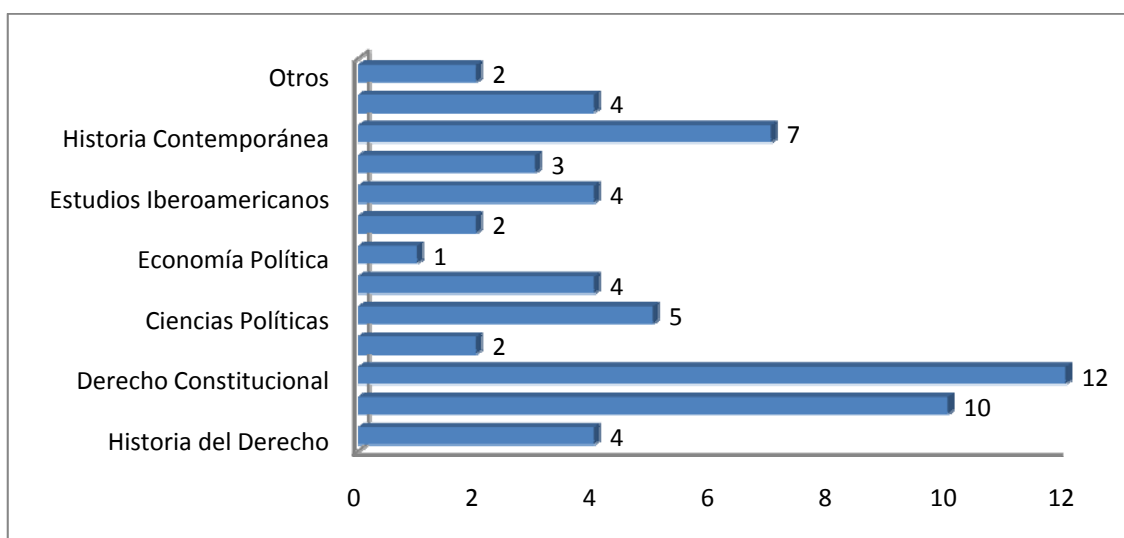
Gráfica 3. Gráfica correspondiente a la comprensión lectora de los usuarios en distintos idiomas. Elaboración propia.

4. En cuanto a la pregunta referida a las áreas de interés, hemos considerado aquellas con las que el CEPC cuenta con publicaciones periódicas, al entender que el centro se concibe especializado en las mismas. Asimismo, su política de incorporación de fondos a su colección está orientada a las actividades y cursos que allí se imparten, con lo cual hemos tratado de sintetizar las materias de los mismos en un listado general de áreas. Este listado se lo hemos mostrado a nuestros usuarios para conocer el dominio de interés que guía sus demandas informativas tanto a través del catálogo como del conjunto de repositorios que integran la colección digital accesible desde el sitio web del CEPC. La lista con los

resultados de las áreas a las que se adscriben nuestros usuarios la mostramos a continuación con su correspondiente gráfica.

Historia del Derecho	4	15%
Derecho Civil	10	37%
Derecho Constitucional	12	44%
Administración Pública	2	7%
Ciencias Políticas	5	19%
Derecho Internacional	4	15%
Economía Política	1	4%
Estudios Africanos	2	7%
Estudios Iberoamericanos	4	15%
Sociología	3	11%
Historia Contemporánea	7	26%
Derecho Comunitario	4	15%
Otros	2	7%

Tabla 14. Dominios de interés para los usuarios encuestados. Elaboración propia.



Gráfica 4. Gráfica correspondiente a las áreas a las que se adscriben los usuarios del CEPC de acuerdo con sus demandas informativas. Elaboración propia.

Como se infiere de la tabla y de la gráfica contamos con tres grandes grupos de áreas de interés: I) El constituido por las áreas de derecho civil y constitucional. II) El constituido

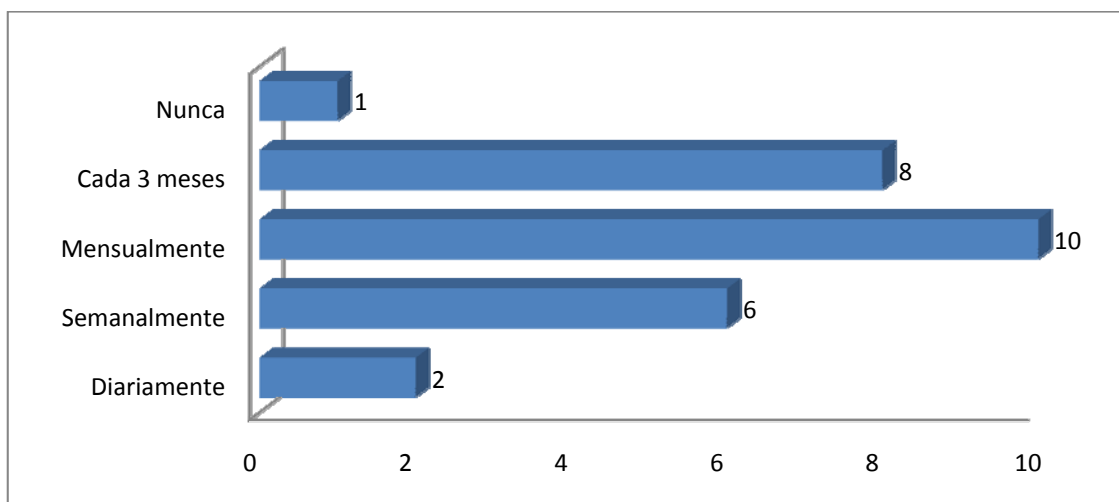
por las áreas de Historia Contemporánea y Ciencias Políticas. III) Resto de áreas. Es normal que las áreas de más interés sean las correspondientes al derecho y más concretamente al derecho constitucional y civil, dado que el CEPC es un referente investigador por sus fondos y actividades en las mismas. Asimismo, es muy empleado por los propios funcionarios del centro por su carácter asesor para el gobierno de la legalidad de sus anteproyectos de ley. Respecto a la Historia Contemporánea, la Historia del Derecho y las Ciencias Políticas, también cuenta con una importante comunidad de usuarios que acuden para la consulta de los importantes fondos antiguos con los que cuenta la biblioteca en temas de historia sociopolítica del viejo régimen. En lo referente al último grupo, se constata una dispersión de áreas de interés que van desde la sociología hasta los estudios africanos⁸¹⁸, o los estudios iberoamericanos, la administración pública o la economía política.

5. En esta quinta pregunta hemos querido mostrar la regularidad con la cual nuestra comunidad de usuarios acude físicamente a la biblioteca para empleo tanto físico como digital de los recursos que se les pone a su disposición. Como vemos en la tabla, a pesar de que muchos de los encuestados son alumnos del propio centro declaran acudir con poca regularidad a la biblioteca. Fundamentalmente acuden cada tres meses o mensualmente. Sin embargo, debemos destacar la existencia de un importante grupo de usuarios que acuden semanalmente, estos son los investigadores propios del centro y funcionarios. Compararemos a continuación estos resultados con los obtenidos de la contestación de la siguiente pregunta acerca de la frecuencia con la que acceden a los recursos digitales del centro desde el sitio web del CEPC.

Diariamente	2	7%
Semanalmente	6	22%
Mensualmente	10	37%
Cada 3 meses	8	30%
Nunca	1	4%

Tabla 15. Periodicidad de acceso a la biblioteca del CEPC. Elaboración propia.

⁸¹⁸ A finales de los años 40 el CEPC editaba una revista de estudios africanos para el análisis de cuestiones políticas relativas a las colonias que España tenía en África. Sin embargo, con la pérdida de Guinea Ecuatorial estos estudios declinaron y se cerró la revista a mediados de los años 50.



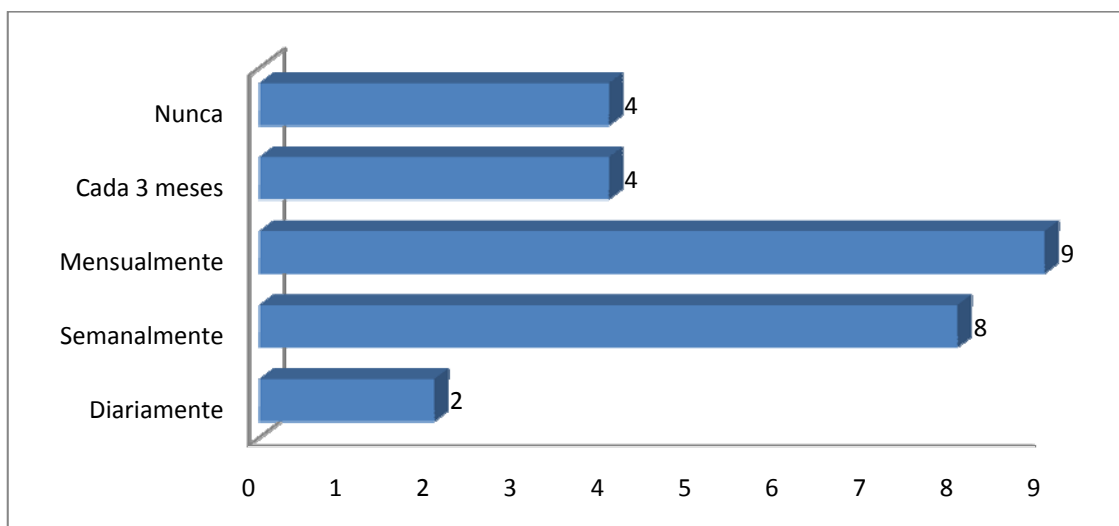
Gráfica 5. Gráfica correspondiente a la frecuencia con la que los usuarios acuden físicamente a la biblioteca del CEPC. Elaboración propia.

6. Como comentamos esta pregunta está en relación con la quinta. En la tabla y en la gráfica vemos cómo no existen diferencias importantes entre el acceso físico y digital a los recursos de la biblioteca. No obstante, existen algunos que vamos a comentar porque nos permitirán introducir el papel que los *topic maps* ejercen en el diseño de la biblioteca digital semántica, así como comprender la importante mejora que se hace de su uso.

Análogamente al caso anterior son muchos los usuarios que acceden mensualmente a la biblioteca digital, pero ahora vemos cómo existen un número análogo de usuarios que acceden semanalmente frente a quienes lo hacían cada 3 meses para la biblioteca física del CEPC. Otra distinción es el elevado número de usuarios que nunca han accedido al sitio web de la biblioteca digital, o que lo hacen cada tres meses. Esto arroja luz sobre uno de los primeros problemas de la biblioteca digital y es que parece haber un grupo de usuarios capaces de realizar un empleo regular de la biblioteca digital frente a otro grupo de usuarios prácticamente incapaces o sin interés por hacerlo. Trataremos con las siguientes preguntas de mostrar algunas causas de ello.

Diariamente	2	7%
Semanalmente	8	30%
Mensualmente	9	33%
Cada 3 meses	4	15%
Nunca	4	15%

Tabla 16. Periodicidad de acceso a la biblioteca digital del CEPC. Elaboración propia.

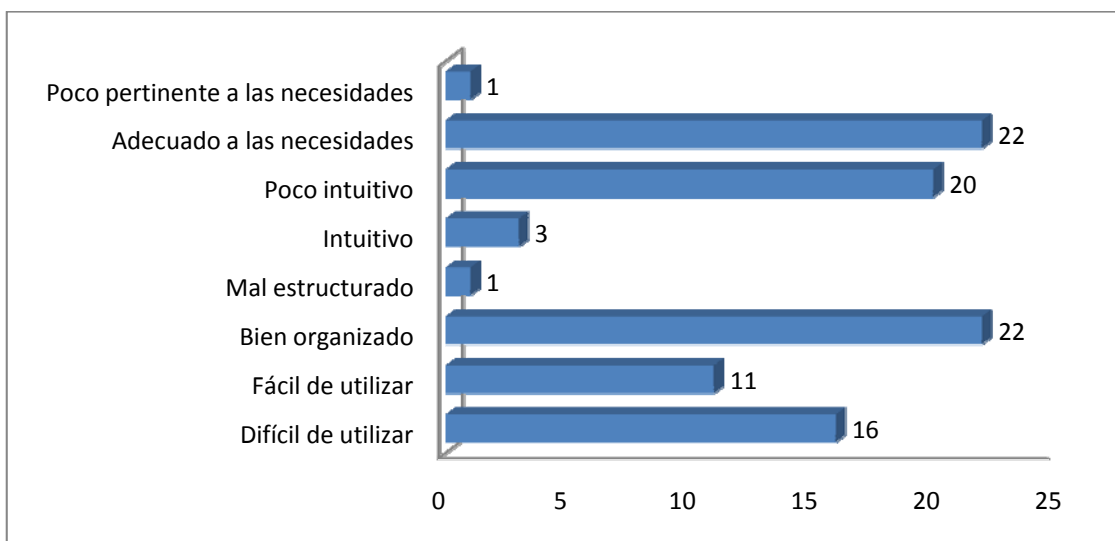


Gráfica 6. Correspondiente a la frecuencia de acceso a la biblioteca digital del CEPC. Elaboración propia.

7. Para conocer un poco más acerca de los dos grupos de usuarios de la biblioteca digital les hemos preguntado acerca de su opinión personal acerca de su impresión general del sitio web. De los datos ofrecidos por la opinión de su comunidad de usuarios se desprende que estamos ante un sitio web que a pesar de estar bien organizado y muy adecuado a sus necesidades, resulta difícil de utilizar y poco intuitivo. Entonces, ¿por qué una biblioteca digital, a pesar de estar bien organizada resulta difícil de utilizar y es poco intuitiva? Trataremos de buscar la respuesta en las siguientes preguntas realizadas a nuestros usuarios.

Difícil de utilizar	16	59%
Fácil de utilizar	11	41%
Bien organizado	22	81%
Mal estructurado	1	4%
Intuitivo	3	11%
Poco intuitivo	20	74%
Adecuado a las necesidades	22	81%
Poco pertinente a las necesidades	1	4%

Tabla 17. Respuesta de los usuarios a la pregunta sobre facilidad de uso de la biblioteca. Elaboración propia.



Gráfica 7. Correspondiente a la facilidad que encuentran los usuarios para el acceso a la biblioteca digital del CEPC. Elaboración propia.

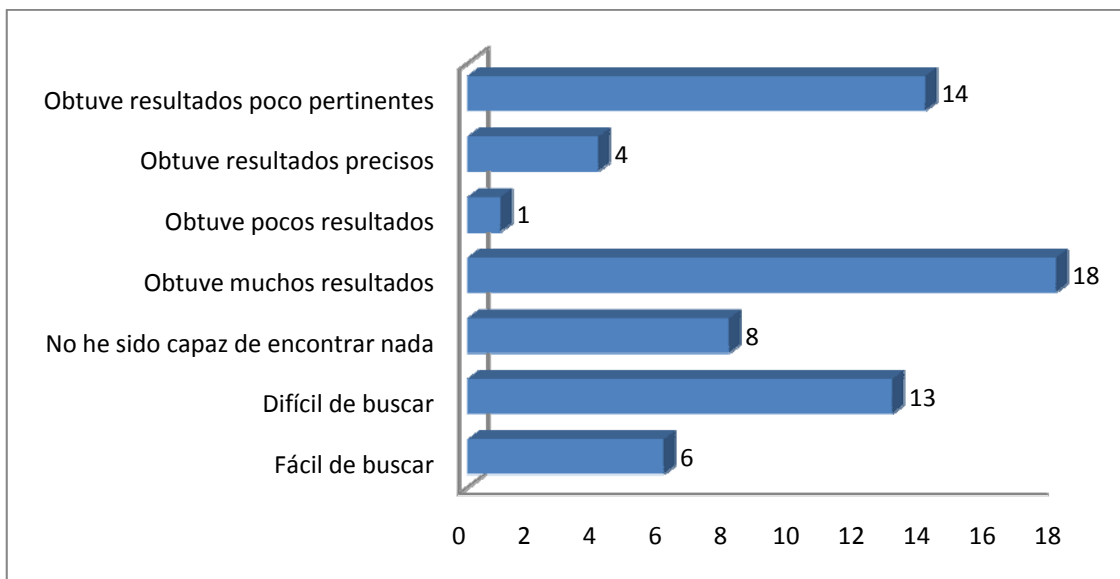
8. Para conocer un poco más acerca de las habilidades de nuestros usuarios con la biblioteca digital les pedimos realizar una pequeña prueba de búsqueda. Se trataba de que buscaran artículos sobre la materia del Senado en el área del Derecho Constitucional. Se pretendía que mostraran su conocimiento sobre las bases de datos y las revistas digitales, concretamente la Revista Española de Derecho Constitucional editada por el propio centro. Esto nos permitirá orientarnos acerca tanto del grado de conocimiento de los recursos de la biblioteca como sobre todo la regularidad real con la que trabajan, no tanto “en” la biblioteca sino “con” la biblioteca.

De los datos que a continuación se muestran se desprende que para la mayoría de los usuarios ha resultado difícil la realización de una búsqueda para la que existen múltiples vías. Sin embargo, los recursos de la biblioteca han operado bien ya que les han ofrecido muchos recursos sobre lo que se les pedía. Asimismo, debemos destacar el elevado porcentaje de usuarios incapaces de realizar la búsqueda, así como la de aquellos usuarios que no fueron capaces de localizar documentos sobre lo que se les pedía. Y, ¿ocurre lo mismo en lo referente a sus habilidades para buscar en el catálogo, tanto si se efectúa en el propio centro como fuera de él, a través del sitio web?

Fácil de buscar	6	22%
Difícil de buscar	13	48%
No he sido capaz de encontrar nada	8	30%

Obtuve muchos resultados	18	67%
Obtuve pocos resultados	1	4%
Obtuve resultados precisos	4	15%
Obtuve resultados poco pertinentes	14	52%

Tabla 18. Respuesta de los usuarios sobre el grado de dificultad de la búsqueda en la biblioteca digital del CEPC que se les pidió realizar. Elaboración propia.



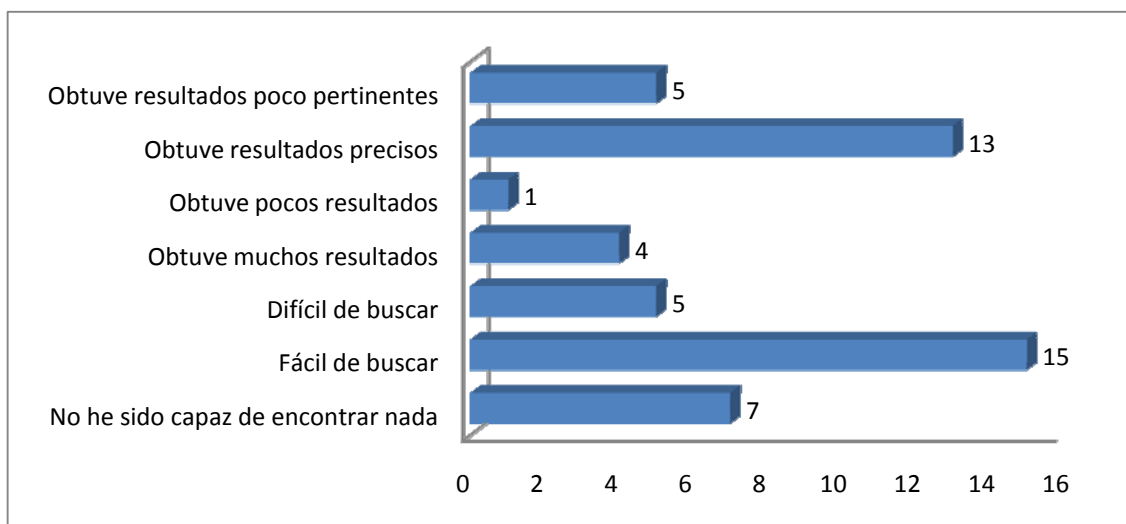
Gráfica 8. Correspondiente al grado de dificultad de las búsquedas en la biblioteca digital del CEPC planteadas al usuario. Elaboración propia.

9. Respecto al catálogo, vamos a ver si efectivamente lo utilizan con regularidad y correctamente por medio de un sencillo ejercicio de búsqueda en el mismo. El ejercicio consiste en localizar una obra de Álvarez Junco sobre la figura del político de la II República Alejandro Lerroux. Hemos constatado que en general no han tenido ningún problema para utilizar el catálogo. Los propios usuarios, tal como vemos en la tabla de datos de a continuación, nos indican que la búsqueda es fácil para localizar con precisión lo que se les pide. No obstante, se manifiesta un pequeño grupo de usuarios que no han podido encontrar la obra que se les pedía o que les ha parecido difícil. Así pues, incluso respecto al catálogo existe una clara cesura entre los usuarios con más o menos habilidad para realizar búsquedas de documentos y aquellos que carecen de habilidades para ello.

No he sido capaz de encontrar nada	7	26%
Fácil de buscar	15	56%
Difícil de buscar	5	19%

Obtuve muchos resultados	4	15%
Obtuve pocos resultados	1	4%
Obtuve resultados precisos	13	48%
Obtuve resultados poco pertinentes	5	19%

Tabla 19. Respuesta de los usuarios sobre el grado de dificultad de la búsqueda que se les pidió realizar en el catálogo. Elaboración propia.



Gráfica 9. Correspondiente al grado de dificultad de las búsquedas en el catálogo planteadas al usuario. Elaboración propia.

6.4. Propuesta de metodología para la edición de un *Topic map* de Biblioteca Digital Semántica Jurídica.

En este último capítulo de la tesis, una vez visto en profundidad lo que significan los *topic maps* como proyección de los futuros modelos semánticos de información, vamos a plantear, una vez identificados algunos aspectos acerca de nuestra comunidad de usuarios cuantitativamente a través de las encuestas y cualitativamente a través de la realización de ejercicios de búsqueda que nos permitieron conocer los hábitos de navegación de nuestros usuarios. Asimismo, hemos partido de ejemplos de bibliotecas digitales semánticas diseñadas con *topic maps*, como la New Zealand Electronic Text Centre⁸¹⁹ o el Tax Map de la agencia tributaria norteamericana⁸²⁰ para explicar el desarrollo de nuestra propuesta metodológica, con la cual hemos construido la biblioteca digital semántica del CEPC.

⁸¹⁹ Biblioteca digital, construida con *topic maps*, New Zealand Electronic Text Centre. En: <http://www.nzetc.org/>. Consultado el 15/02/2010.

⁸²⁰ IRS Tax Map. En: <http://taxmap.ntis.gov/taxmap/>. Consultado el 15/02/2010.

La metodología propuesta está destinada a clarificar el proceso global de diseño del *topic map*, los sujetos implicados y las principales cuestiones a tener en cuenta cuando se edita uno. Abarca dos etapas, una primera que comprende la parte inicial del proyecto hasta su publicación, y una segunda que es la alimentación y mantenimiento del mismo. Todo ello pretende ser una propuesta para la elaboración de un *topic map* que cubra el ciclo de vida de la biblioteca digital semántica; sin embargo, quedan aspectos por desarrollar, fundamentalmente a causa de la actual situación de normalización de distintos aspectos de los *topic maps* como son su interoperabilidad con los *schemas* de metadatos o los lenguajes específicos de consulta y definición de inferencias⁸²¹.

Hoy día, existe una amplia literatura acerca del modelado de datos, parte de la cual es muy relevante para el desarrollo de ontologías *topic map*. Entre estos modelos contamos con UML, el modelo de Chen o el modelo ORM, con los que ya existe una literatura científica para el modelado de *topic maps* presentada en los congresos TMRA en Leipzig⁸²². Por otra parte, aunque el modelado de datos es diferente del modelado de ontologías, ya contamos con varios trabajos sobre ello, pero no así respecto al modelado de *topic maps*, al que no le son aplicables directamente los principios de modelado de ontologías. En diseño de *topic maps*, los trabajos más relevantes fueron publicados por Ahmed⁸²³, especialmente el referido a modelos para diseño de ontologías, así como los publicados por Garshol⁸²⁴ o Passin⁸²⁵.

Una de las primeras recopilaciones de metodologías para el desarrollo de ontologías, cuando todavía no se había aprobado la norma 13250 para *topic maps*, es el trabajo de Visser⁸²⁶, pero no hemos encontrado que ninguna de las metodologías existentes para el desarrollo de ontologías sean aplicables de forma directa al desarrollo de *topic maps* o para su aplicación a bibliotecas digitales semánticas. No obstante, muchas de ellas fueron utilizadas como contribución a la metodología que nos planteamos para construir el *topic map* del CEPC, tanto durante el proceso como en la redacción de las normas de diseño del

⁸²¹ Cfr. 5.2.

⁸²² GULBRANDSEN, A. Conceptual modeling of Topic maps with ORM versus UML. En *TMRA 2005*, Lecture Notes in Computer Science, vol. 3873. Leipzig : Springer, 2006. En: http://dx.doi.org/10.1007/11676904_8. Consultado el 15/02/2010.

⁸²³ AHMED, K. Topic map design patterns for information architecture. En *XML 2003*, IDEAlliance. En: <http://www.techquila.com/tmsinia.html>. Consultado el 15/02/2010.

⁸²⁴ GARSHOL, L. Towards a methodology for developing Topic maps Ontologies. En MAITCHER, L. ; SIGEL, A. ; GARSHOL, L. *Leveraging the Semantics of Topic maps*. Berlin : Springer, 2007.

⁸²⁵ PASSIN, T. *Explorer's Guide to the Semantic Web*. Greenwich : Manning, 2004.

⁸²⁶ JONES, D.M. ; BENCH-CAPON, T.J.M. ; VISSER, P.R.S. Methodologies for Ontology development. En *Proceedings IT&KNOWS Conference, XV IFIP World Computer Congress*. Budapest, 1998.

mismo. La metodología más destacada con la que contamos es DILIGENT⁸²⁷, y el tutorial para su modelado que más utilizamos fue el propuesto por Noy y McGuinness⁸²⁸.

6.4.1. La ontología del *Topic map*.

El término “ontología” del que ya hemos proporcionado varias definiciones, vamos a definirlo para nuestra metodología como una colección de *topics*, *names*, *occurrences*, *associations* y *role types* empleados en un *topic map* particular o una colección de *topic maps* estructurados de forma análoga⁸²⁹.

No estamos hablando de una ontología en el sentido de una base de conocimiento de un dominio particular, destinado a proporcionar inferencia lógica; sino que, el tipo de ontología para *topic maps* es un modelo conceptual para un dominio particular, que hace empleo de una estructura y terminología familiar a los usuarios finales⁸³⁰. De ahí, que el campo jurídico sea propicio para su empleo en el diseño de bibliotecas digitales semánticas jurídicas⁸³¹. Esto explica el surgimiento en los últimos años de iniciativas en cuanto a ontologías jurídicas⁸³², pero no tanto respecto a ontologías para *topic maps* jurídicos. Sin embargo, existen ejemplos como el Tax Map, cuya ontología vemos comparativamente a grandes rasgos con nuestro *topic map* para el CEPC.

⁸²⁷ VRANDECIC, D. ; PINTO, S. ; TEMPICH, C. ; SURE, Y. The DILIGENT knowledge processes. *Journal of Knowledge Management*, vol.9, nº 5, 2005.

En: http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2005_kmjournal_diligent.pdf. Consultado el 15/02/2010.

⁸²⁸ NOY, N. F. ; MCGUINNES, D. *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford : Stanford University, 2001. En: <http://www-ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf>. Consultado el 15/02/2010.

⁸²⁹ GARSHOL, L. Towards a Methodology for Developing *Topic Maps* Ontologies. En MAITCHER, L. ; SIGEL, A. ; GARSHOL, L. *Leveraging the Semantics of Topic Maps*. Berlin : Springer, 2007. p.21.

⁸³⁰ Id.

⁸³¹ BREUKER, J. ; VALENTE, J. ; WINKELS, R. Use and reuse of legal ontologies in Knowledge Engineering and Information Management. *Law and the Semantic Web*. Berlin : Springer, 2005. p.37.

⁸³² Cfr. 3.4.3.

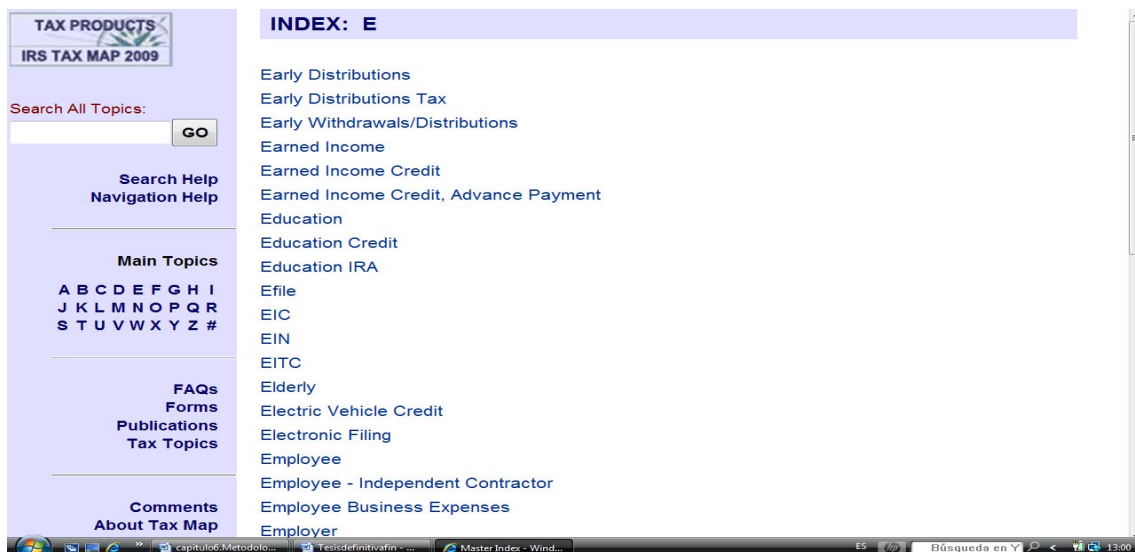


Figura 85. Clasificación de *topic types* en Tax Map.

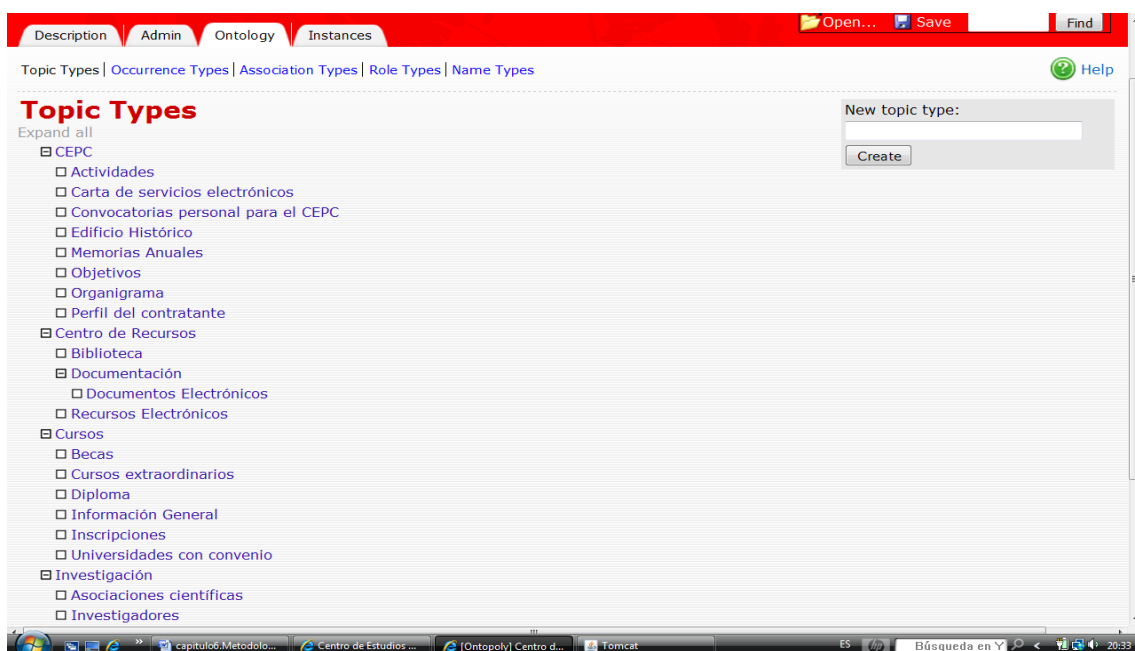


Figura 86. Clasificación de *topic types* en el *topic map* del CEPC. Elaboración propia.

La ontología ha sido crucial para el desarrollo de nuestra biblioteca digital semántica organizada con *topic maps*; hasta el punto de poder decir que esta determina cómo colocar todos los elementos en la biblioteca digital semántica. Entre las razones para ello tenemos que⁸³³:

⁸³³ YANG, H. ; LEE, Ch. Building *Topic Maps* using a text mining approach. En Zong, N. et al. (eds.). *ISMIS 2003*. Berlin : Springer, 2003. pp.310ss.

- 1) Define la estructura del *topic map* usada, análogamente a como lo haría un modelo entidad-relación para una base de datos relacional. Por tanto, la ontología determinó lo que contiene el *topic map*.
- 2) La estructura de la biblioteca digital semántica necesariamente sigue la estructura de la ontología de manera estrecha; de tal forma que se puede afirmar que la ontología es el proyecto que se sigue para definir la arquitectura de información del mismo.
- 3) Todo el código XTM de la aplicación se escribió de acuerdo con la ontología. Esta incluye el nivel de visualización y el interfaz de administración en omnigator con el cual integrar datos.

El diseño de una ontología *topic map* requiere de un análisis analítico y abstracto de la realidad que se requiere representar, así como un interfaz capaz de recoger los requisitos de un amplio abanico de usuarios para poder lograr un consenso en las decisiones que se tomen en el diseño de la ontología entre la comunidad de usuarios a la que está destinada. Por tanto, las opiniones que hemos recogido con las entrevistas, tanto a los usuarios como a los profesionales del centro han sido esenciales para el diseño del *topic map* del CEPC.

En general, podemos decir que hemos imaginado la ontología *topic map* en tres partes:

1º) Un procedimiento de desarrollo de la ontología a través de la representación en un mapa conceptual de los principales elementos a considerar en la biblioteca digital semántica para la ontología *topic map*. En este proceso representacional se detalla cómo aproximarse al desarrollo del *topic map* con un mapa conceptual diseñado con el editor The Brain⁸³⁴, que nos permite relacionar los principales *topics* y aportar información adicional de cada uno de ellos, tanto por parte nuestra como por parte de todas las personas a quienes consultamos sobre cómo relacionar los elementos *topic* para mejorar la navegación en el futuro *topic map* del CEPC. El procedimiento está orientado a ayudar al diseñador a asegurarse que su desarrollo se adecúe a los requisitos de construcción de una biblioteca digital semántica; así como, asegurar un consenso global de los usuarios y los técnicos de la biblioteca acerca de la ontología diseñada para facilitar su futuro mantenimiento y actualización. Asimismo, el tipo de visualización que permite, árbol hiperbólico, es la misma que utilizamos para visualizar el *topic map* del CEPC, Vizigator⁸³⁵, del editor Ontopia.

⁸³⁴ THE BRAIN. En: <http://www.thebrain.com/>. Consultado el 15/02/2010.

⁸³⁵ Vizigator para el CEPC. En: http://localhost:8080/omnigator/plugins/viz/viz.jsp?tm=centro_de_estudios_pol_ticos_y_constitucionales.xtm&id=id409&redirect=%2Fomnigator%2Fmodels%2Ftopic_complete.jsp%3Ftm%3Dcentro_de_estudios_pol_ticos_y_constitucionales.xtm%26id%3Did409. Consultado el 15/02/2010.

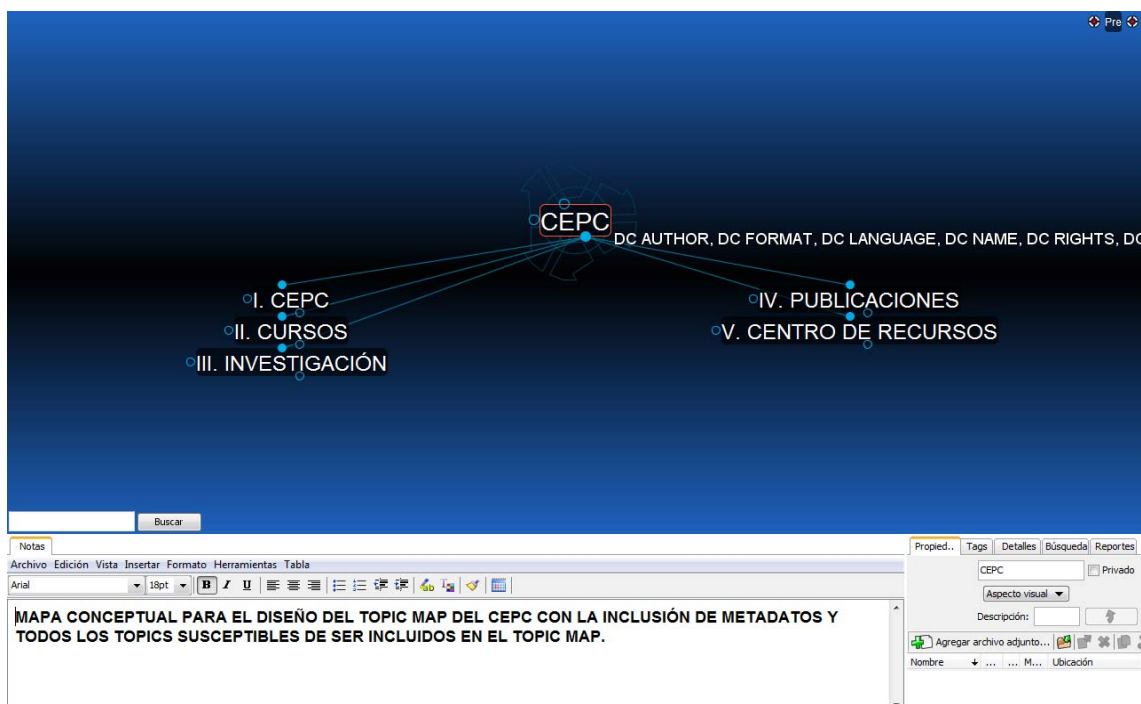


Figura 87. Mapa conceptual para el diseño de la ontología *topic map* del CEPC. Elaboración propia.

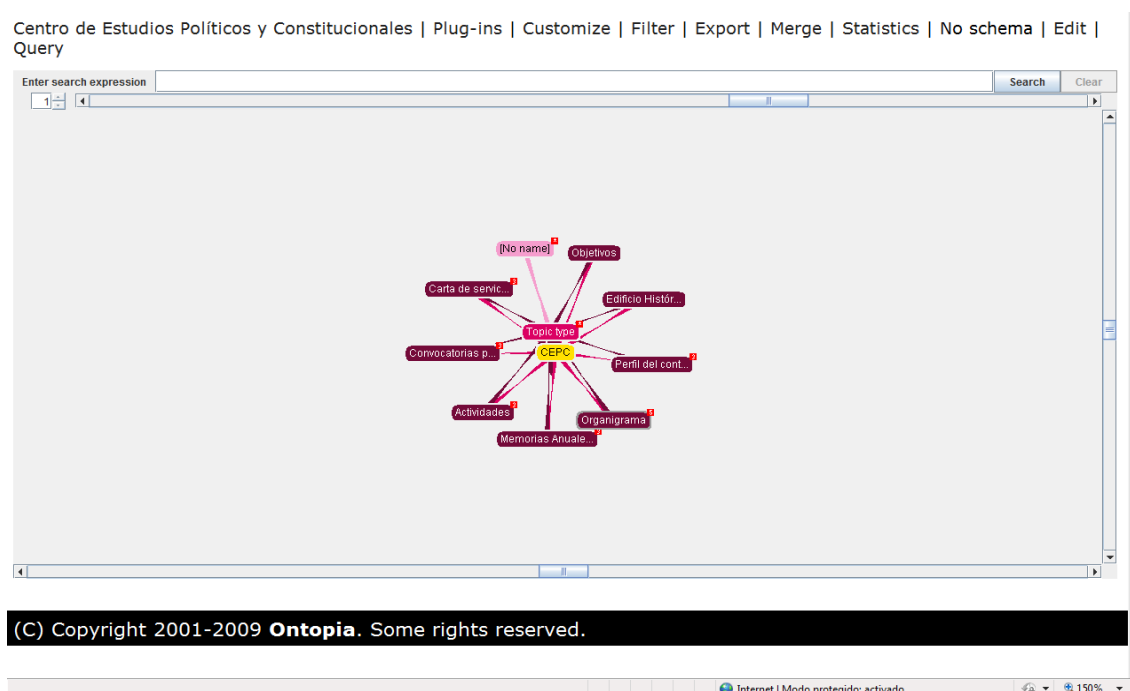


Figura 88. Visualización del *topic map* del CEPC con Vizigator. Elaboración propia.

2º) Un conjunto de guías para el desarrollo de ontologías susceptibles de ser aplicadas al diseño coherente y acorde con unas reglas de *topic maps*. Esta serie de guías para la correcta construcción de las ontologías nos permiten prescindir para un *topic map* reducido como el nuestro del lenguaje TMCL con el cual se deberían definir las reglas que gobiernan su

organización. Sin embargo, partimos de guías para ontologías y lenguajes documentales que nos aportan un conjunto de reglas comunes a los *topic maps*, como puede ser la necesidad de una función para cada término cuando se establece una relación con otro. Entre estas guías contamos con la guía de diseño de ontologías de Protégé⁸³⁶, o la guía para el diseño de tesauros de Gilchrist⁸³⁷ y Vanda Broughton⁸³⁸, o la guía de *topic maps* de Garshol⁸³⁹.

3º) Una biblioteca de modelos o casos prácticos de aplicación de ontologías para el diseño de *topic maps* que organicen bibliotecas digitales semánticas. El objetivo es recoger soluciones proporcionadas en el diseño de este tipo de bibliotecas a problemas comunes surgidos durante el diseño de nuestra biblioteca. Con ello, contamos con una ayuda para resolver problemas complicados de diseño, de modo que las soluciones sean consistentes con los mejores casos prácticos aplicados. Fundamentalmente, nuestra biblioteca de modelos ha estado integrada por la biblioteca digital del NZETC⁸⁴⁰, los *topic maps* de tesauros diseñados por Pepper⁸⁴¹ y el *topic map* Tax Map de Biezunski⁸⁴².

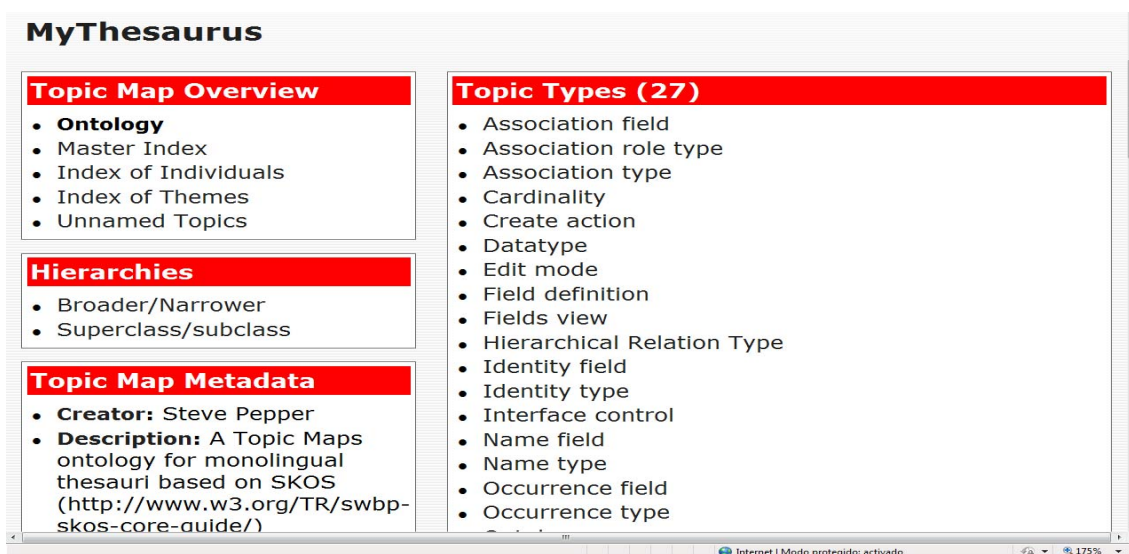


Figura 89. *Topic map* de tesauros diseñado por Pepper.

⁸³⁶ NOY, N. ; MCGUINNES, D. *Ontology Development 101: A Guide to Creating Your First Ontology*. En: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf. Consultado el 15/02/2010.

⁸³⁷ AITCHISON, J. ; GILCHRIST, A. ; BAWDEN, D. *Thesaurus construction and use: a practical manual*. London : Europa Publications, 2000.

⁸³⁸ BROUGHTON, V. *Essential thesaurus construction*. London : Facet, 2006.

⁸³⁹ GARSHOL, L. *Towards a Methodology for Developing Topic Maps Ontologies*. Op. cit.

⁸⁴⁰ NEW ZEALAND ELECTRONIC TEXT CENTRE. En: <http://www.nzetc.org/>. Consultado el 15/02/2010.

⁸⁴¹ PEPPER, S. *My Thesaurus*. En: http://localhost:8080/omnigator/models/topicmap_complete.jsp?tm=MyThesaurus.xtm. Consultado el 15/02/2010.

⁸⁴² BIEZUNSKI, M. *The internal revenue service Tax Map*. En: www.coolheads.com/MBPUBS/xml/europe2003/1rstm4.ppt. Consultado el 15/02/2010.

6.4.2. El procedimiento de desarrollo del *Topic map*.

6.4.2.1 Personal necesario para la edición y diseño del *Topic map*.

Los proyectos de bibliotecas digitales semánticas suelen ser iniciativas de organizaciones que desean poner en marcha un proyecto de biblioteca con un abanico amplio de recursos de información complejos, y con una serie compleja de funciones para satisfacer las demandas de sus usuarios⁸⁴³. Como hemos comprobado en la realización de esta tesis, si bien el diseño lo ha hecho enteramente el autor con ayuda de los usuarios y personal de biblioteca, los distintos participantes susceptibles de formar parte del proyecto juegan distintos roles que pueden contribuir al diseño del *topic map* desde sus propias necesidades y demandas. Es por ello que la interacción entre el rol de cada uno de ellos y el *topic map* se hace esencial para la formulación de una metodología de diseño.

Los roles que hemos definido para nuestra metodología son⁸⁴⁴:

- I) Un gestor de proyecto. Una persona o personas responsables de la dirección y gestión del proyecto.
- II) Administradores de sitios web. El grupo de personas que tienen la última palabra en la toma de decisiones que tengan que ver con la biblioteca digital, especialmente en lo relacionado con la asignación de funciones en el portal a otras personas, con excepción de los proveedores de las fuentes de información.
- III) Editores. Personas responsables de todos aquellos aspectos relacionados con la comunicación entre la biblioteca digital y sus usuarios, incluidos el diseño de sitio, la arquitectura de información, la terminología y la actualización de contenidos.
- IV) Modelador de la ontología. Quien crea la ontología *topic map* para el portal.
- V) Desarrolladores. Quienes hacen efectiva la biblioteca digital semántica incluyendo el interfaz de presentación, el sistema editorial y la integración de datos.

⁸⁴³ FERRER SAPENA, A. ; PESET MANCEBO, F. ; MORENO NÚÑEZ, M. ; LLORET ROMERO, N. *Guía metodológica para la implantación de una biblioteca digital universitaria*. Gijón : Trea, 2005, p.43-44.

⁸⁴⁴ GARSHOL, L. *Towards a Methodology for Developing Topic Maps Ontologies*. Op. cit., p.22.

VI) Propietarios de sistemas autónomos para el acceso a datos. Los responsables de los sistemas que proporcionan acceso al portal a través de la integración de datos.

VII) Diseñador de interactividad. Persona responsable de la definición de la estructura de las páginas desde el punto de vista de su visualización e interactividad con el portal.

VIII) Usuario final. Personas que utilizan la biblioteca digital operativa para satisfacer sus demandas informativas.

IX) Autores. Quienes escriben texto para la biblioteca, o añaden contenidos al *topic map*. Este rol puede ser el que jueguen también los editores.

X) Experto en el dominio del Derecho, la Ciencia Política o la Sociología. Aquella persona conocedora del dominio modelado. Esta misma función también puede ser, e incluso debería ser realizada conjunta y coordinadamente por un editor y el propietario del sistema autónomo.

Es posible que algunos de los miembros participantes del proyecto jueguen más de uno de los roles mencionados para un mismo proyecto de *topic map*. Todos los roles, a excepción de los administradores web y el personal editorial, pueden ser satisfechos con personas ajenas a la organización responsable de la biblioteca digital semántica.

6.4.2.2. El proceso.

El proceso se define por una secuencia de fases, cada una de las cuales implica una subcolección de roles con su correspondiente colección específica de elementos que pasarán a integrar el *topic map*. El proceso no tiene por objetivo explicar en sí mismo la totalidad del desarrollo de la biblioteca digital semántica, tan sólo se refiere a la ontología *topic map*, con lo cual se omiten muchos pasos necesarios para la puesta en funcionamiento final de la biblioteca digital semántica.

En el siguiente gráfico se muestra el esquema global de todas y cada una de las fases de desarrollo de la ontología *topic map* que explicaremos seguidamente⁸⁴⁵.

⁸⁴⁵ Ibid., p.23.

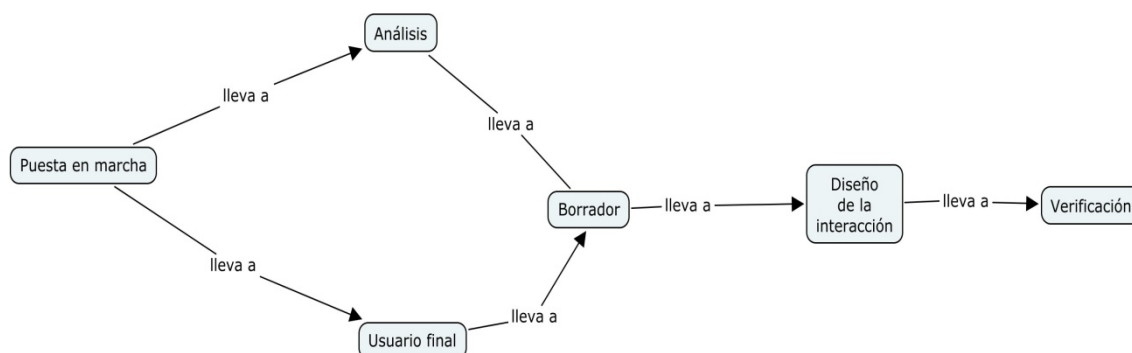


Figura 90. Esquema de las fases de desarrollo de la ontología *topic map* según la metodología de esta tesis.

I) Fase de puesta en marcha. Su propósito es ayudarnos en la comprensión del proyecto, de forma que seamos capaces de aprender a concebir una visión final del mismo, el equipo necesario para ponerlo en marcha, la comunidad de usuarios finales, y cualesquiera sujetos externos con los cuales la biblioteca pueda llegar a necesitar algún tipo de relación. Resulta esencial que seamos capaces de alcanzar una visión integradora de los recursos de información, así como de los sistemas de información con los que contará nuestra biblioteca digital semántica del CEPC para que estos puedan ser interoperables. Durante esta fase contamos como diseñadores con el asesoramiento de los gestores de la biblioteca y servicio de publicaciones, los administradores web y los editores. El formato de esta fase toma la forma de entrevistas permanentes donde los miembros del CEPC, tanto usuarios como trabajadores, van presentando sugerencias sucesivas del mismo, seguido por entrevistas con miembros de otros proyectos relevantes como puedan ser proyectos de implantación de repositorios. El resultado de esta fase fueron un conjunto de notas que nos ayudaron a obtener una visión global del proyecto, así como proporcionar unos primeros criterios de selección de las fuentes informativas con las que la biblioteca digital semántica va a contar. Estas notas las añadimos al mapa conceptual de partida que vimos durante el proceso de desarrollo de la ontología *topic map* y que construimos con The Brain.

II) Fase de usuario final. El propósito de esta fase es conocer las demandas informativas de nuestra comunidad de usuarios a la que está destinada la biblioteca digital semántica; y si es posible, qué terminología emplean para realizar sus consultas en relación con los objetos de información que puedan resolver sus demandas. Para ello, nos hemos basado en el trabajo de evaluación de *topic maps* de Myongho Yi⁸⁴⁶. Esto nos permitirá asegurarnos que la

⁸⁴⁶ YI, M. *Topic Maps-based Ontology and Semantic Web*. Saarbrücken : VDM Verlag, 2008.

ontología *topic map* empleada por la biblioteca cuenta con una buena interactividad con los usuarios finales y les proporcione la información más pertinente posible a sus demandas. Esta fase se concretó mediante las preguntas de competencia y las posteriores entrevistas a los usuarios, así como las encuestas.

Para recoger las respuestas de los usuarios a las preguntas de competencia les pasamos en la biblioteca una tarjeta con la pregunta que se muestra a continuación y se les pedía que la resolvieran en un tiempo máximo de diez minutos tal y como se muestra a continuación.

Instrucciones y pregunta: Por favor, acceda a la siguiente dirección: www.cepc.es	Indique tres obras sobre el concepto de Justicia en Kelsen e indique su estrategia de búsqueda, los recursos del CEPC empleados y por qué ha empleado éstos.
Tiempo de búsqueda	Comienzo del ejercicio a: 10h14'
Respuesta	He localizado las siguientes obras: 1) Normas jurídicas y análisis lógico; 2) ¿Qué es la justicia?; 3) La garantía jurisdiccional de la Constitución. He ido al catálogo y en el apartado de cualquier campo he puesto: "kelsen justicia". El recurso que he empleado es el catálogo del CEPC porque es el recurso que me da más resultados.
Tiempo de búsqueda	Finalizado a: 10h23'

Tabla 20. Tarjeta de evaluación cualitativa de los recursos del CEPC. Diseño basado en el trabajo de Myongho Yi.

III) Fase de análisis. En el CEPC, realizamos únicamente tres entrevistas de las cuales obtuvimos un conjunto de notas, no sólo como mostramos en la tabla acerca de la resolución a la pregunta planteada, sino también sobre las impresiones que nos ofreció el resultado de la entrevista. Estas impresiones corresponden a la fase de análisis y las integrábamos en el mapa conceptual. Asimismo, en esta fase de análisis integramos los comentarios procedentes del análisis de resultados de la encuesta.

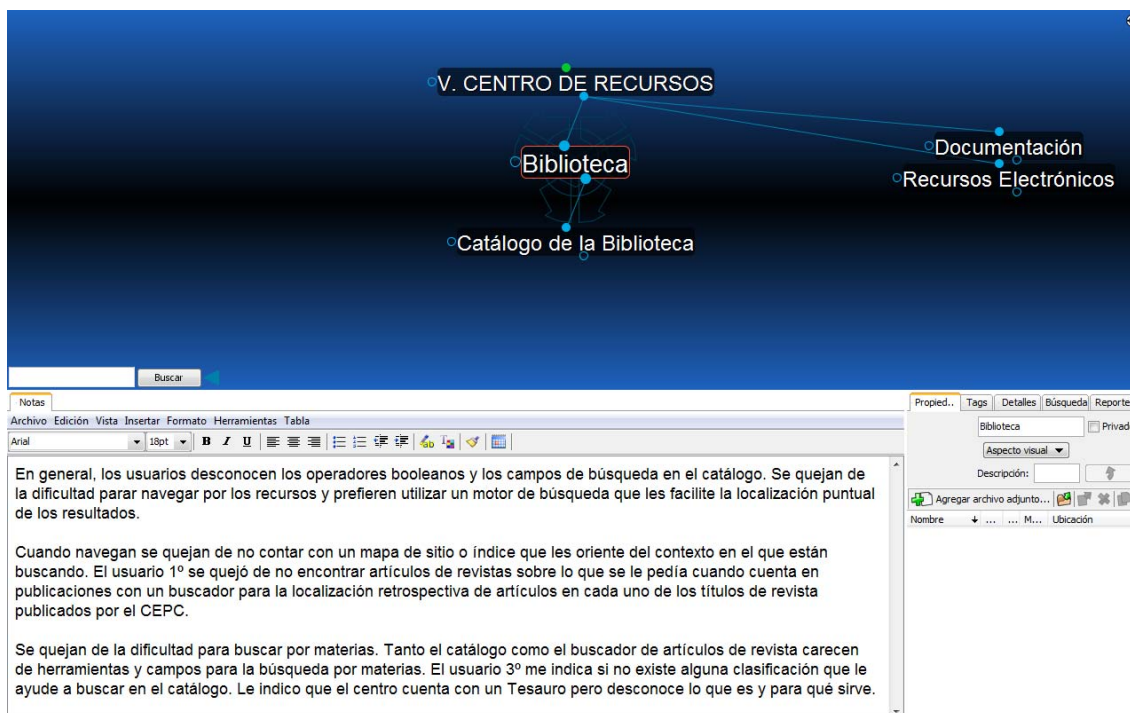


Figura 91. Ejemplo de notas recogidas en el mapa conceptual. Elaboración propia.

Los comentarios añadidos al mapa conceptual fueron usados en una posterior fase de verificación. La finalidad de esta fase es permitirnos lograr una comprensión profunda de los datos a partir de las fuentes identificadas en la fase anterior. Para las fuentes ya existentes, debemos intentar conseguir documentación, esquemas de información, y si es posible exportación de datos y estudios sobre los mismos. Para las fuentes todavía no establecidas, debe contentarse con cualquier información que pueda obtenerse. Esta fase implica primeramente interactuar con los propietarios de las fuentes.

El formato para esta fase es un seminario con los editores, el administrador web y algunos expertos en el campo del derecho. Pueden incluirse entrevistas con las editoriales propietarias de los recursos de información, junto con un análisis de los mismos que

responde a la investigación surgida de las respuestas proporcionadas por las editoriales durante la evaluación de la colección.

El resultado de esta fase será un borrador del flujo de información en la biblioteca digital entre los distintos niveles que la estructuran. Así se consiguen notas referidas a la ontología que muestran los principales *topics* y *association types* de cada una de las fuentes informativas.

IV) Fase de borrador. El propósito de esta fase es crear una imagen inicial de cómo debe parecer la ontología, establecer una visión compartida del contenido de la biblioteca entre el modelador y las personas que influyen en el funcionamiento de la biblioteca digital semántica. El borrador no es un proyecto ni cerrado ni corregido en esta etapa; lo fundamental, es que las personas con capacidad de influencia en el diseño de la biblioteca, sean capaces de juzgar si hay o no importantes omisiones o malinterpretaciones en el borrador. En esta fase están implicados el modelador, el gestor del proyecto, los desarrolladores y los editores.

Esta fase normalmente toma la forma de un borrador creado por el modelador que envía a las personas que participan en el funcionamiento de la biblioteca para que lo revisen. Esta labor permite revelar vacíos entre la información disponible en la biblioteca y las necesidades de los usuarios. En ese caso, puede ser necesario revisar el trabajo de las primeras etapas, o bien extender el ámbito del proyecto. Puede haber muchos ciclos como éstos de revisión y correcciones antes de que un borrador final llegue a ser aprobado. En el peor de los casos puede que las dos primeras fases deban ser repetidas, por lo que se hace recomendable tener entrevistas con los principales actores de la biblioteca para presentarles personalmente los borradores y explicárselos, y evitar así tener que repetir las fases de redacción del borrador final. Podremos así, obtener como resultado de la fase un documento descriptivo de la ontología.

V) Fase de diseño de la interacción. La finalidad de esta fase es consensuar un interfaz para la biblioteca y asegurarse de que la ontología se pueda representar en ese interfaz. El diseño de la interacción de la biblioteca digital y la ontología están íntimamente relacionados, por lo que en esta fase es cuando se refina la ontología. En ella toman parte el gestor del proyecto, el modelador, el diseñador de la interacción, el personal editorial y los desarrolladores. El formato de la fase son seminarios en los que toman parte el gestor del

proyecto, el personal editorial, el modelador de la ontología y el diseñador de la interacción. En los seminarios, se presentan una serie de esbozos del interfaz que representan las distintas partes de la biblioteca digital semántica. El modelador de la ontología participa en los seminarios para asegurarse que el diseño de la interacción es coherente con la ontología, y que se modifique el borrador de la ontología sólo en aquellos aspectos estrictamente necesarios. Después de los seminarios, el modelador actualiza el borrador de la ontología y se produce el primer *topic map* dirigido por la ontología diseñada y de acuerdo con un interfaz ajustado a las características y requerimientos de la misma. El resultado de esta fase es la documentación del diseño de la interacción, una versión actualizada del borrador de la ontología y un primer *topic map*. *Topic map* que puede estar bien almacenado con una sintaxis propia como XTM o LTM, o bien guardado en una aplicación para la edición de *topic maps*.

VI) Fase de verificación. En esta fase se verifica que la ontología corresponde a la información de las fuentes que constituyen la colección digital de la biblioteca, que satisface las demandas informativas de los usuarios y satisface las exigencias de los administradores de sitio del CEPC. Además, se incluye la verificación de que la calidad de la información es la adecuada para los objetivos de la biblioteca digital. En esta fase las modificaciones de la ontología son menores, fundamentalmente afectan al diseño de la interacción. Quienes están implicados en su desarrollo son el modelador, las editoriales proveedoras de recursos informativos, el gestor del proyecto y los editores.

Esta fase toma la forma del resultado del trabajo del modelador con los proveedores de objetos de información y los contextos informativos asociados. Esto supone que los editores deben verificar la ontología frente a los resultados de la fase de usuario final y cualquier requisito exigido en la documentación. El resultado serán versiones actualizadas de la documentación, así como una versión actualizada del *topic map*. También se incluirá una versión inicial del código de conversión y un borrador del *topic map* con los datos convertidos. Una vez se haya completado la fase de verificación será cuando pueda comenzarse el desarrollo de la biblioteca digital semántica.

6.4.2.3. La documentación de la ontología.

El *topic map* que estructura la biblioteca digital semántica es el resultado final del modelado de la ontología sobre la cual existe una dependencia física, pero la documentación de la ontología es, en muchos sentidos, igualmente importante. Esta transmite la filosofía del

CEPC que tanto la dirección del centro como sus trabajadores pretenden reflejar en el funcionamiento de la biblioteca digital semántica a través del *topic map* que la estructura. La documentación, consiste en un único mapa conceptual que sirve tanto como introducción como referencia. Comienza con una introducción general destacando la estructura del *topic map*, a continuación se sigue con la definición detallada de cada una de las secciones del mismo. Para ello, hemos dividido el *topic map* en módulos o partes de las que se puede ofrecer una explicación más detallada. Los PSI's y los nombres son definidos e interrelacionados para su posterior referencia.

La documentación más detallada se presenta en diagramas de clases UML que representan de manera más detallada el *topic map*. Sin embargo, lo que hemos hecho ha sido dividir los diagramas en partes que sean visualmente manejables pero completos para las partes de la ontología que cubren. Por ello, la documentación para cada parte de la ontología incluye su PSI, una definición, y una descripción del mismo, de forma que dejamos claras las fuentes suministradoras de instancias para cada uno de los *topic types*; así como información para su mantenimiento así como las personas responsables, fechas de actualización, o herramienta empleada para su alimentación. Incluso, se indica qué sintaxis es común a las distintas partes del *topic map*, para que no haya incoherencias en su utilización y asegurar la interoperabilidad, tanto entre los distintos elementos del *topic map* como de cada uno de ellos con los sistemas de gestión de contenidos ya implantados en la biblioteca.

6.4.3. Guía propuesta para el desarrollo del *Topic map*.

La guía se utiliza durante el modelado del *topic map* para asegurar la coherencia en la adopción de decisiones de diseño. En general, la guía asume que el lector está familiarizado con los conceptos básicos de modelado de datos y aquellos aspectos particulares para el modelado que intervienen en el diseño de ontologías para la creación de *topic maps*.

6.4.3.1. Principios generales.

Si bien se pueden proporcionar algunas reglas básicas de diseño de una ontología *topic map*, debemos tener en cuenta que no existen reglas absolutas que no se puedan romper en casos particulares de diseño de este tipo de ontologías, de forma que una de las habilidades básicas de un modelador de *topic maps* es saber cuándo dejar de seguir las recomendaciones. En cualquier caso, en muchas ocasiones se nos han planteado varias alternativas para el

modelado de una misma información y el criterio ha sido seleccionar aquella que más se adecuaba al proceso lógico de consulta por parte de un usuario.

En general, podemos decir que el objetivo fundamental del diseño de *topic maps* es hacer que su ontología sea tan próxima al dominio que se desea representar como sea posible. Es decir, para cada *topic type* y *association type* debe haber en el dominio representado alguna clase de materia con la cual se pueda establecer una correspondencia tan próxima como sea posible.

Otro principio general es que la redundancia debe ser evitada, de tal forma que el diseño del *topic map* no se debe realizar de forma que incluya información susceptible de ser obtenida más que de una única forma. El proceso de filtrado de estructuras redundantes en el *topic map* es en el fondo lo mismo que la normalización en el modelo relacional de bases de datos, ya que las convenciones son en muchos casos las mismas. Esto quiere decir que por norma general, los constructos con un contexto implícito deben ser evitados.

6.4.3.2. Topic Types.

Los *topic types* son el eje vertebrador de nuestra ontología *topic map*, y por ello el punto de partida en el proceso de modelado. La cuestión ha estado en decidir cuándo un término es un *topic type*. Normalmente, un *topic type* es un nombre para un concepto abstracto, una clase o plantilla de cosas en cuyo universo se pueden integrar instancias de ese tipo. Es análogo a una categoría en el sentido de subdivisión de un dominio de conocimiento en subcolecciones nombradas. La diferencia reside en el hecho de que las categorías carecen de instancias. Así, para el caso del portal actual del CEPC, “lenguas” es una categoría pero no un *topic type* dado que carece de instancias. No hay elementos que sean lenguas.

Topic Maps y Tecnologías de la Información Jurídica: un modelo de recuperación de información para bibliotecas digitales semánticas jurídicas.

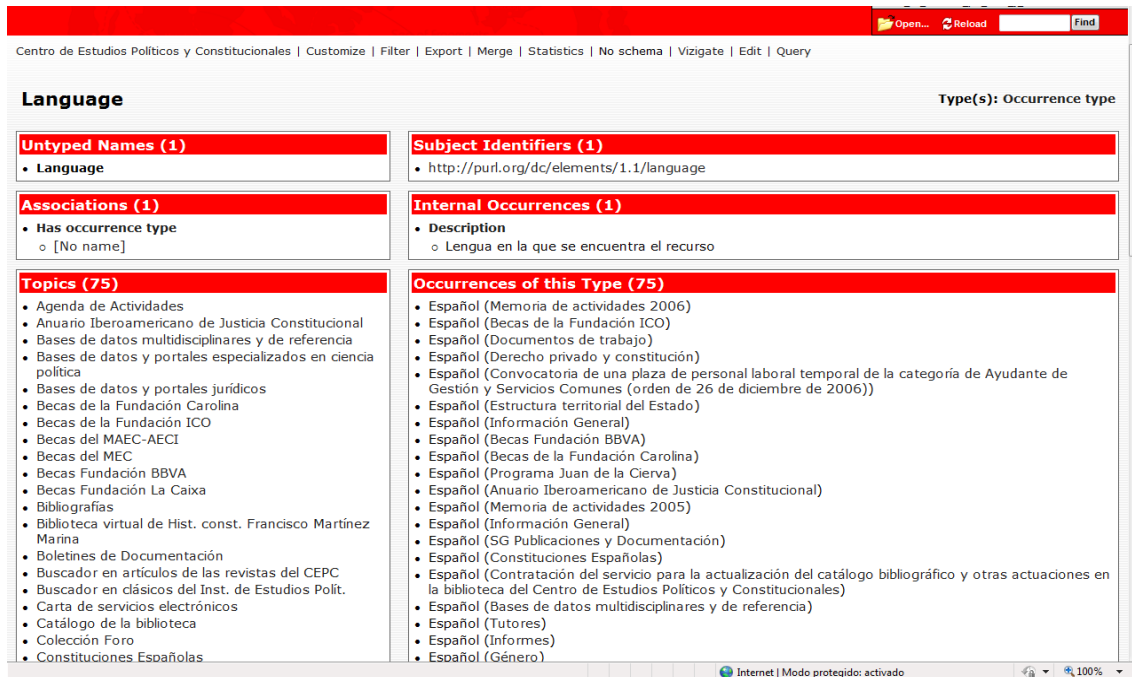


Figura 92. Ejemplo de *topic* que no es *topic type*. Elaboración propia.

En cambio, “Revistas” es un *topic type* porque contiene elementos que son revistas.

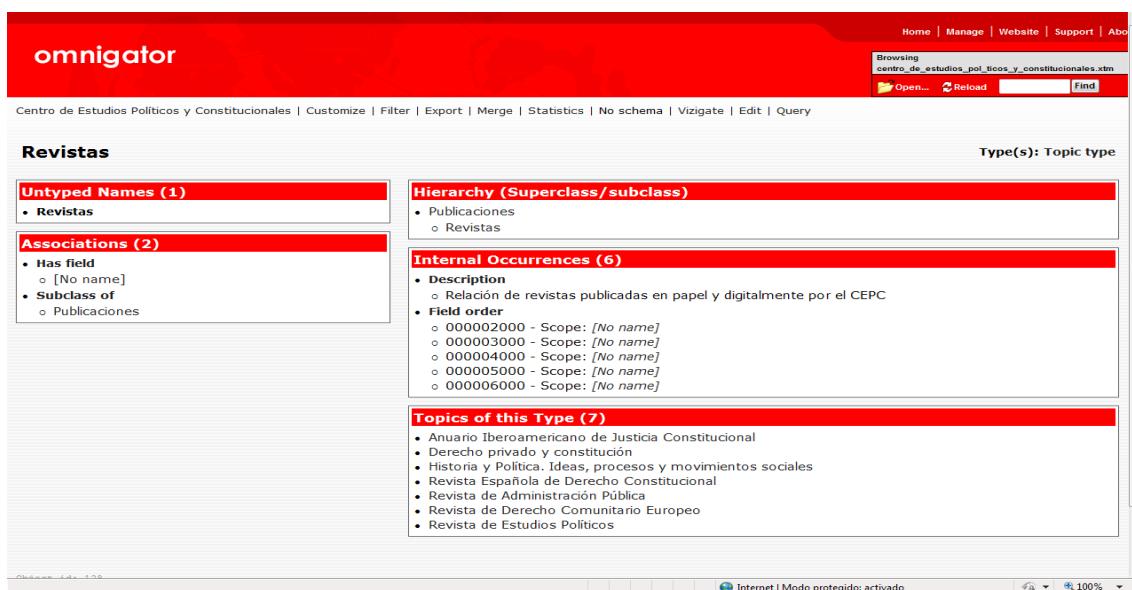


Figura 93. Ejemplo de *topic type*. Elaboración propia.

Las características de un *topic type* se manifiestan mejor a través de la relación entre este y sus instancias. En primer lugar, porque un buen *topic type* es definido cuando se minimizan las dudas sobre que términos se califican como instancias. Normalmente, una instancia, por su naturaleza, es una instancia de un tipo o clase dada. De ello se desprende que cada instancia debe ser instancia de un mismo tipo siempre; y por tanto, dada una instancia,

existe un *topic type* del cual es una instancia. Por ejemplo, consideremos el *topic type* “jurista”. Nos podemos preguntar si en nuestro *topic map* es adecuado considerarlo como *topic type*, en principio parece claro su ámbito; sin embargo, no está en la propia naturaleza de los juristas el serlo. El jurista es una persona con una profesión pero no tiene nada que ver con las etapas vitales de la persona, ni con ámbitos personales ajenos a su profesión de jurista. Por tanto, el jurista es en primer lugar una persona con independencia de cuál sea su labor y su situación, lo que nos lleva a seleccionar “persona” como *topic type* y no “jurista”.

6.4.3.3. Jerarquía de clases.

La norma 13250 define un tipo de *association* “supertipo-subtipo” para la construcción de jerarquías de clases. Estas operan análogamente al proceso de modelado para los tesauros.

Siempre es posible crear una jerarquía de clases que abarque todos los *topic types* de la ontología, y que se elabora en sentido descendente mediante una cadena de subtipos específicos. No obstante, en general, no es recomendable introducir más clases en la jerarquía de clases a no ser que existan unas necesidades específicas.

Algunas de las razones que pueden justificar la introducción de nuevas clases en la colección básica de los *topic types* son:

- I) Con el objeto de aclarar la relación entre los *topic types*, e identificar sus características en común. En el caso de nuestro *topic map* ocurrió con los *topic types* “publicaciones digitales” y “revistas”. Lo que hicimos fue introducir un supertipo común que denominamos “Publicación” que recoge el hecho de que ambos *topic types* son muy similares, proporcionando un nombre a las propiedades conjuntas de ambos.
- II) Con el objeto de simplificar la expresión de las restricciones de la ontología. Normalmente esto significa la creación de un supertipo común que recoge el marco de restricciones compartidas entre sus subtipos. El ejemplo de “publicación” visto ilustra cómo sus dos subtipos tienen casi idénticas propiedades, y la superclase evita tener que repetir la inserción por duplicado de propiedades comunes.

- III) A fin de simplificar la búsqueda en la aplicación. Es probable que nuestros usuarios deseen buscar únicamente por “publicaciones” más que por “publicaciones digitales” o “revistas” que complicaran la búsqueda para el usuario.
- IV) A fin de simplificar la comunicación con los usuarios de la aplicación. Los usuarios pueden llegar a encontrar las subclases con demasiado detalle como para poder comprenderlas, en tanto que las superclases pueden estar representadas por términos mucho más significativos.

6.4.3.4. Names.

Los *names* son más fáciles de concebir si los entendemos como un tipo especial de campo en un *topic type*, la dificultad que ofrecen está en saber cuándo un campo debe ser o no un nombre. En general, hemos tomado por regla que siempre que deseábamos mostrar los contenidos del campo en el interfaz de usuario como etiqueta para un *topic*, entonces ese será un nombre para el *topic*.

Normalmente, todo *topic type* debe tener un nombre por defecto, el cual debe servir como nombre bajo el que se muestra el *topic* salvo en contextos concretos como el cambio de idioma del *topic map*. La cardinalidad es exactamente una ya que si hubiese más de un software no habría modo de escoger entre los nombres en ausencia de un contexto; si no hubiera un nombre por defecto, el *topic* carecería de un nombre por el poder ser nombrado. Es posible añadir tantos nombres adicionales como se deseen de acuerdo con las normas para la adición de *occurrences*.

6.4.3.5. Occurrences types internos.

Los *occurrences types* internos son equivalentes a las propiedades o atributos en otros paradigmas de modelado, y por tanto no presentan ningún tipo de problema de modelado para los *topic maps*. Así, típicos *occurrence types* internos para el *topic type* “organigrama” en nuestro *topic map* pueden ser: “cargo o responsabilidad en el centro”, “dirección de correo electrónico”. Como en otros paradigmas de modelado para cada *occurrence type* se debe definir la cardinalidad y los *data type*.

6.4.3.6. Occurrence types externos.

Formalmente hablando, los *occurrences types* externos son tan sólo unos tipos de *occurrence* cuyos tipos de datos son URI. Sin embargo, son conceptualmente especiales al presentar unas cualidades específicas. En parte, vienen del hecho que las URI's almacenadas en el *topic map* referencian recursos de información externos, lo cual constituye una entidad en sí misma, pero no llega a ser un *topic* del *topic map*.

Existen dos maneras de representar los documentos en un *topic map*:

- I) Empleo de *occurrences* externas. En este caso ninguna otra información acerca del documento puede ser recogida más que la clase y contexto de su relación con uno o varios de los *topics* del *topic map*.
- II) Creación de un *topic* para el documento y empleo de *associations* en vez de *occurrences* externas. En este caso, no hay limitaciones sobre lo que pueda ser dicho acerca del documento.

Todo esto se puede resumir en una regla general y es que: si deseamos decir algo acerca de un documento, creamos un *topic* para él. Si no deseamos decir nada usamos una *occurrence* externa. Sin embargo, existen otras convenciones en el diseño de *topic maps* acerca del empleo de *occurrences* externas, tales como limitaciones donde pueden ser almacenadas, o los formatos que puedan emplearse. En cualquier caso, todo ello debiera estar documentado en la documentación acerca del *topic map* si no fuera posible recogerlo en un esquema formal.

6.4.3.7. Association Types.

En *topic maps*, existe un único modo de expresar relaciones entre *topics*: con *associations*. Esto hace que el diseño de las *associations* sea relativamente simple dado que su definición es similar a como se realiza en el modelo de entidad-relación de Chen o en UML; sin embargo, la diferencia respecto a estos modelos reside en que un mismo *association type* puede repetirse en distintas partes del *topic map*.

La principal dificultad en el diseño de *association types* reside en el tratamiento que se le da a los roles. Los *topic maps* proporcionan *associations* de cardinalidad arbitraria, lo que ha supuesto muchos problemas para los diseñadores de *topic maps*.

Es por ello, que hemos hecho uso, para el diseño del *topic map* del portal CEPC, de las siguientes reglas para la aplicación de *associations*:

- I) Todas las *associations* del mismo tipo deben contar con el mismo juego de *role types*. En algunos casos, se permite que algunos de los *role types* sean opcionales, pero su omisión se hace por la información que falta.
- II) Cada *association* representa una relación del mundo real.
- III) Los *association types* deben, normalmente, tener la cardinalidad más baja sin que suponga una pérdida de información.
- IV) Cada *role type* no podrá aparecer más de una vez en cada *association*.

Un caso especial que ocurre en ocasiones es el referido a la relación de simetría, donde los dos *topics* de una *association* participan del mismo modo en la relación. Por ejemplo, tenemos el caso de la relación simétrica que se establece en el apartado de “Publicaciones” entre “Revistas digitales” y “Publicaciones digitales”. La relación es la misma por parte de ambos *topics* ya que si “Revista digital” está relacionada con “Publicación digital” entonces necesariamente “Publicación digital” está relacionada con “Revista digital”. En *topic maps* esto se representa por medio de una *association* del tipo “Formato” con dos roles que son el mismo: “digital”.

Para las *association types* los elementos a definir son los siguientes:

- I) Los *role types* en *associations* de ese mismo *type*. Para cada *role type*, los *topic types* donde actúan los roles deben ser definidos, así como su cardinalidad dentro de la *association*. Normalmente cardinalidad 1 a 1.
- II) La cardinalidad de cada combinación *role type/association type* para cada uno de los *topic types* que juegan un *role type* dado en su correspondiente *association type*.

6.4.3.8. Identificadores.

En general, todos los *types* en el *topic map* tienen definidos sus correspondientes PSI's, lo que permite simplificar enormemente la identificación de los *types* en cuanto a código y la configuración de archivos de distintas clases. Además, permite la simplificación de la integración de datos que vayan siendo introducidos, lo que junto a la creación de un indicador de materia para cada *type* resulta bastante útil para que pueda ser comprensible tanto por nuestros usuarios como por parte del personal de la biblioteca; obligando al diseñador del *topic map* a considerar en detalle la semántica de cada uno de los *types*.

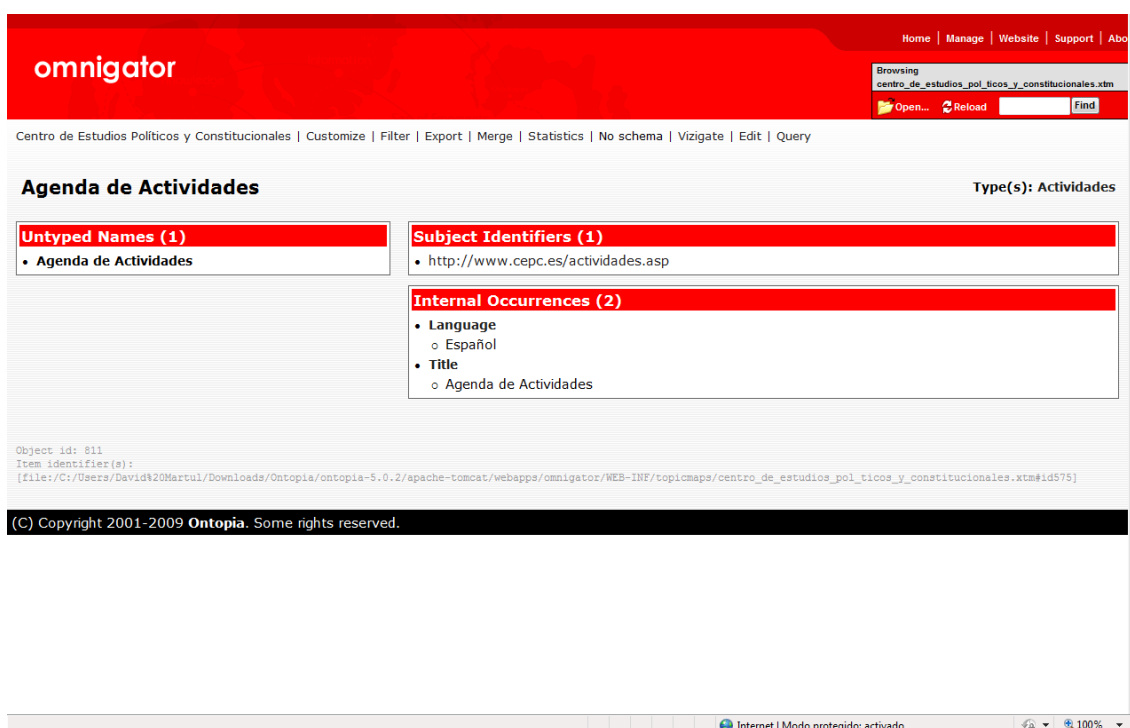


Figura 94. Ejemplo de PSI para un *topic type*. Elaboración propia.

Como regla general se aconseja la reutilización de *types* definidos en otros lugares donde haya correspondencia entre la semántica de los *types*, así como definir los *types* restantes en un PSI *namespace* individual. Esto significaría una URI que termina en “/”, donde los PSI's de los *types* individuales en el *namespace* de cada uno tiene URI's que comparten un prefijo común y sólo difieren en la última parte de una URI que no debe tener ni barras ni espacios vacíos.

6.5. Evaluación por parte de los usuarios de la biblioteca digital semántica del CEPC con *topic maps*.

6.5.1. Marco general de la evaluación final del *topic map*.

Una vez realizado el *topic map* para una biblioteca digital semántica jurídica del CEPC, pasamos al último punto, la evaluación por parte de nuestros usuarios del mismo en relación con sus habilidades de búsqueda y demandas informativas. Sin embargo, para centrar los resultados de esta evaluación vamos a tratar de contextualizarlos un poco más.

Los artículos de revistas, informes técnicos, actas de congresos, sitios web, diccionarios, tesauros, libros, tesis y manuales son algunos de los distintos tipos documentales manejados en la biblioteca del Centro de Estudios Políticos y Constitucionales⁸⁴⁷ (CEPC), con la doble finalidad de apoyar la investigación y el aprendizaje o formación de los usuarios que acuden a este centro en calidad de alumnos de los cursos en Derecho Constitucional y Ciencia Política⁸⁴⁸. Además de su variada tipología documental, el servicio de biblioteca cuenta con un elevado número de publicaciones que cada año se incorporan a la colección⁸⁴⁹, tratándose de un crecimiento exponencial en muchos de sus campos, especialmente de aquellos en los que sus publicaciones se realizan en formato electrónico. En cuanto a tipologías documentales, esto se manifiesta en revistas, aunque cada vez es más frecuente editar actas de congresos en formato digital. A ello se añade que hoy día, ya es posible acceder a la edición facsímil de las ediciones en papel de sus revistas, lo cual proporciona acceso digital tanto a sus ediciones antiguas como a las actuales. Esta labor de digitalización de la colección, orientada a proporcionar un elevado nivel de difusión no sólo para las revistas editadas por el CEPC sino de toda su colección, resulta de especial importancia para las áreas de más rápido desarrollo en el ámbito jurídico. Por otro lado, la biblioteca del CEPC cuenta con tipos documentales como los diccionarios, enciclopedias, tesis y repertorios legislativos que por el gran volumen de información que abarcan les resulta especialmente adecuado el formato digital. Este confiere una serie de ventajas en términos de accesibilidad, dinamismo y posibilidad de enlazarse con fuentes informativas externas a través de los hiperenlaces. Otra razón es la posibilidad de ahorrar costes de impresión, pero sobre todo son la facilidad de distribución, acceso y difusión las razones

⁸⁴⁷ Véase las secciones de Biblioteca, Documentación y Publicaciones. En: <http://www.cepc.es/>. Consultado el 10/02/2010.

⁸⁴⁸ REAL DECRETO 1269/1997, de 24 de julio, por el que se regula la organización y funcionamiento del Centro de Estudios Políticos y Constitucionales. En: http://www.cepc.es/real_decreto.asp. Consultado el 10/02/2010.

⁸⁴⁹ Boletín de información bibliográfica. En: http://www.cepc.es/inform_biblio.asp. Consultado el 10/02/2010.

que hacen este formato más adecuado para una comunidad de usuarios especializada en el ámbito jurídico. Todas ellas son suficientes para justificar la necesidad de comenzar a organizar una colección digital que integre servicios de valor añadido, tales como web social o la recuperación por contexto, mediante la implantación de una biblioteca digital semántica para el CEPC⁸⁵⁰.

La colección del CEPC, siguiendo la tendencia de otros centros de documentación jurídicos, está proporcionando una diversidad de repositorios digitales jurídicos, como subject gateways, portales, bibliotecas digitales, bases de datos. Cada uno de ellos con sus propias interfaces gráficas de usuario, herramientas de navegación y recuperación de información. Sin embargo, no existen muchos trabajos en el campo de la documentación jurídica sobre cómo presentar un interfaz común a todos los repositorios de una biblioteca⁸⁵¹. Es este uno de los campos al que busca dar solución la biblioteca digital semántica jurídica, ya que no sólo han cambiado la colección y servicios de la biblioteca del CEPC, sino también la interacción de los usuarios con la misma. Ahora, la estrategia más común empleada por ellos es “localizar y leer” en lugar de “leer y localizar” típico del entorno papel⁸⁵². Tenopir y King estudiaron los hábitos de lectura y observaron que la cantidad de publicaciones disponibles en línea ha cambiado el modo en que los investigadores las leen. Un hallazgo de su estudio es que actualmente los investigadores tienden a leer más artículos que antes, pero el modo en que los leen ha cambiado. De hecho tienden a leer más artículos pero emplean de media menos tiempo en su lectura, hasta el punto que en pocas ocasiones leen un artículo entero sino que hojean de forma superficial buena parte del texto y sólo en algunas partes hacen una lectura muy detenida. Por otra parte, el tiempo de lectura y búsqueda de cada artículo se ha duplicado porque además de tener que buscar los artículos también se tiene la necesidad de buscar materias dentro de cada uno de los artículos⁸⁵³.

En cuanto a la interacción de los usuarios con los documentos, la biblioteca jurídica del CEPC contiene miles de páginas a través de las cuales se puede navegar y realizar

⁸⁵⁰ KRUK, S. *Semantic Digital Libraries. Improving usability of information discovery with semantic and social services*. Lulu.com, 2010. p.11

⁸⁵¹ KRUK, S. ; MCDANIEL, B. *Semantic Digital Libraries*. Berlin : Springer-Verlag, 2009, p.219-220.

⁸⁵² TENOPIR, C. ; KING, D.W. Reading Behaviour and Electronic Journals. *Learned Publishing*, 15, 2002, pp.262-263.

⁸⁵³ KING, D. ; MONTGOMERY, C. H. After migration to an electronic journal collection. Impact on Faculty and Doctoral students. *D-Lib Magazine*, vol.8, nº12, 2002. En: <http://www.dlib.org/dlib/december02/king/12king.html>. Consultado el 15/02/2010.

búsquedas por texto libre. Esto es posible por la digitalización llevada a cabo de las revistas editadas por el propio CEPC⁸⁵⁴, donde cada archivo cuenta con dos niveles o capas superpuestas:

- i) una capa superior con la imagen o facsímil del ejemplar correspondiente de la revista.
- ii) otra inferior con el texto reconocido. Esto permite seleccionar y copiar texto, así como realizar búsquedas por contenido.

The screenshot displays the website of the Centro de Estudios Políticos y Constitucionales (CEPC). On the left, a navigation menu lists options: 'Revistas', 'Información general', 'Presentación', 'Consejo de Redacción/Asesor', 'Normas para los autores', and 'Compra de número'. The main content area features a header for 'Revista de Estudios Políticos > NÚMERO' and a section for 'número 147, Enero/Marzo 2010'. Below this, a table provides a detailed table of contents, organized into sections: 'Cubiertas, Portadillas y Sumario', 'Artículos', 'Notas', and 'Recensiones'. Each section lists the author, title, and a brief description of the article. For example, under 'Artículos', there are entries by Diego Palacios Cerezales, Kamal Mejahdi, Francisco Martínez Mesa, Inmaculada Szmolka Vida, and Kerman Calvo. The 'Notas' section includes an entry by Eguzkí Urteaga. The 'Recensiones' section lists reviews by Mario Kölling, Irene Sobrino Gujarró, and María José Villaverde. The website footer includes the copyright notice '©2008 Cepc. Pza. de la Marina Española 9. 28071. Madrid-España. T. 91 540 19 50' and a browser address bar showing 'http://revistas.cepc.es/revistas.aspx?IDR=3&IDN=687&IDA=27691'.

Autor	Título
Cubiertas, Portadillas y Sumario	
	Cubiertas, Portadillas y Sumario del número 147.
Artículos	
DIEGO PALACIOS CEREZALES	«Extraños cuerpos políticos»: el nacimiento del movimiento social en el Portugal del siglo XIX.
KAMAL MEJAHDI	La evolución democrática de los islamistas marroquíes: incentivos y condiciones.
FRANCISCO MARTÍNEZ MESA	Entre la utopía y la necesidad: una reflexión sobre el cosmopolitismo sansimoniano.
INMACULADA SZMOLKA VIDA	Los regímenes políticos híbridos: democracias y autoritarismos con adjetivos. Su conceptualización, categorización y operacionalización dentro de la tipología de regímenes políticos.
KERMAN CALVO	Movimientos Sociales, y reconocimiento de derechos civiles: la legalización del matrimonio entre personas del mismo sexo en España.
Notas	
EGUZKI URTEAGA	Los sindicatos en Francia.
Recensiones	
por Mario Kölling	JULIO BAQUERO CRUZ y CARLOS CLOSA MONTERO (Eds.): European Integration from Rome to Berlin: 1957-2007.
por Irene Sobrino Gujarró	ABRAHAM BARRERO ORTEGA y MANUEL TEROL BECERRA (Coords.): La libertad religiosa en el Estado social.
por Inmaculada Szmolka Vida	FERRÁN IZQUIERDO BRICHS (Ed.): Poder y regímenes en el Mundo Árabe contemporáneo.
por María José Villaverde	FÉLIX OVEJERO: Incluso un pueblo de demonios: democracia, liberalismo, republicanism.

Figura 95. Estructura en dos niveles de acceso a los artículos a texto completo de la Revista de Estudios Políticos editada por el CEPC.

Las características generales de este sistema que nos han servido como punto de partida para la biblioteca digital semántica del CEPC por su similitud con el concepto de los *topic maps* son⁸⁵⁵:

- 1) Cada revista se estructura en una colección de números, clasificados por años.
- 2) Cada número de la revista se desglosa en una serie de artículos, clasificados por secciones.
- 3) Cada artículo de la revista está digitalizado en un archivo en formato pdf.

⁸⁵⁴ REVISTA DE ESTUDIOS POLÍTICOS. En: <http://revistas.cepc.es/revistas.aspx?IDR=3&IDN=687>. Consultado el 15/02/2010.

⁸⁵⁵ CEPC publicaciones. En: http://www.cepc.es/Inicio_libros.asp. Consultado el 15/02/2010.

- 4) El sistema permite realizar búsquedas selectivas. Introducir texto y establecer criterios de búsqueda por revista, título, autor, sección e intervalo de números o de años. También se puede acceder directamente a un número de una revista.
- 5) Una vez accedido a un artículo, existe la posibilidad de mostrar u ocultar el índice de navegación. Este permite tanto la navegación a través de los artículos de la revista, como la navegación a través de todos los números de la revista.

Hemos solicitado a nuestros usuarios la realización de consultas en las revistas digitales, observando el esfuerzo que dedicaban a la recuperación desde diferentes perspectivas y objetivos para el diseño definitivo del *topic map*. Bases de datos como Aranzadi Westlaw, que indiza una gran cantidad de revistas científicas, como los portales jurídicos Iustel o Tirant on-line tienen el mismo objetivo de proporcionar acceso a la literatura jurídica en formato digital. Sin embargo, en el entorno de las colecciones digitales se proporciona la posibilidad de ir más allá del nivel documental para entrar el ámbito de las materias. Razón ésta por la cual, aunque la estrategia del “localizar luego leer” no sea la más adecuada para todos los tipos de textos⁸⁵⁶, es importante ser capaz de apoyarla, particularmente cuando se haga uso de herramientas de referencia tales como enciclopedias, diccionarios o repertorios jurídicos. Por ello, pensamos que los *topic maps* se adecúan mejor a una biblioteca digital que los tradicionales sistemas de búsqueda por operadores, mas orientados al concepto de “leer luego localizar”. De hecho, hasta el momento se han intentado efectuar algunas aproximaciones, a nivel documental, para mejorar la navegabilidad interna en los textos por medio de una visualización apropiada pero sin resultados satisfactorios⁸⁵⁷. Estos problemas se pretendieron solventar con el empleo de lenguajes de marcado para textos jurídicos⁸⁵⁸, con los cuales los usuarios pudieran localizar información sobre la base de las estructuras documentales. Sin embargo, lo que el usuario demanda es poder efectuar un acceso a los textos conducido por *topics*; es decir, la posibilidad de indicarle los extractos de texto específicos que contienen un *topic* sobre el que se habla. Un *topic* es cualquier materia tratada en una parte concreta de un texto. En este sentido difiere del acceso conducido por palabras clave en considerar al *topic* no sólo como una simple palabra clave sino como

⁸⁵⁶ TENOPIR, C. ; KING, D.W. Reading Behaviour and Electronic Journals. *Learned Publishing*, 15, 2002. p.265.

⁸⁵⁷ HEARST, M. TileBars: Visualization of term distribution information in full text information access. En *Proceedings of CHI'95*, 1995. En: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.53.538>. Consultado el 15/02/2010.

⁸⁵⁸ NOGALES FLORES, T. ; MARTÍN GALÁN, B. ; ARELLANO PARDO, M.C. Informática, Derecho y Documentación. Experiencias y posibilidades de aplicación de los lenguajes de marcado de texto (SGML, HTML y XML) a los documentos jurídicos. En *Encuentro sobre Informática y Derecho*. Madrid : Instituto de informática jurídica, 2002. En: <http://www.bib.uc3m.es/~bmartin/publicaciones/icade2002.pdf>. Consultado el 15/02/2010.

extracto de documentos. También difiere del acceso proporcionado por los sumarios de los libros en que su objetivo es seleccionar partes más precisas de información. Esta necesidad es común en el ámbito jurídico, dado que muchas tareas de búsqueda necesitan la recuperación de información específica en forma de definiciones o explicaciones de conceptos.

Como resultado de nuestra metodología hemos diseñado un *topic map* que nos ha permitido explorar la posibilidad del acceso, “localizar y leer”, con textos jurídicos, así como facilitar la accesibilidad en línea de aquel material que no ha sido producido para ello a fin de poder proporcionar al lector métodos de acceso que vayan más allá de la descarga o lectura superficial de grandes corpora textuales.

Los usuarios que evaluaron nuestro *topic map* fueron fundamentalmente licenciados que se encuentran realizando algún postgrado en el dominio del Derecho y las Ciencias Políticas, con algunos conocimientos básicos de búsqueda en internet y catálogos, y con conocimientos previos que les dimos sobre esta tecnología. El *topic map* consta de *topics* fundamentalmente extraídos del tesoro del ministerio de presidencia, el tesoro Eurovoc y las materias del sitio web del CEPC. Cada *topic* cuenta con una serie de hiperenlaces a secciones de los distintos textos de la colección coherentes con el *topic*. Los usuarios encuestados, al navegar por el *topic map*, nos indican en la encuesta si logran tener una mejor perspectiva del dominio.

De acuerdo con la encuesta que realizamos al comienzo de la investigación sabíamos que contábamos con dos grupos claros de usuarios. Por un lado, aquellos que eran desconocedores casi por completo tanto del catálogo como de las búsquedas en las bases de datos y buscadores así como desconocedores de los sistemas de navegación con directorios y mapas de sitio. Por otra parte, contamos con un pequeño grupo de usuarios, fundamentalmente trabajadores del CEPC y alumnos del postgrado que allí se imparte.

Como hemos visto, realizamos un análisis cualitativo de los hábitos de búsqueda entre los usuarios de este segundo grupo para diseñar el *topic map*, a quienes además se les explicaba en qué consistía la tecnología de los *topic maps*, especialmente en relación con el potencial de búsqueda que proporciona al usuario. Al final de las consultas se les pedía una cuenta de correo electrónico a donde enviarles un breve cuestionario para que proporcionaran sus

impresiones acerca de los aspectos que más les había interesado de los mismos. De los 27 encuestados inicialmente, se contó con 18 de ellos para las pruebas de análisis cualitativo, a quienes se les había pedido en su momento la realización de una serie de pruebas durante la segunda parte del experimento como parte de la metodología para adaptar el diseño del *topic map* a las habilidades y demandas informativas de los usuarios. El cuestionario final que se les envió a partir de una aplicación denominada Google Docs es el que se muestra en la figura de a continuación.

Gmail Calendar Docs Reader La Web Más ▾ dgmartul@gmail.com | Mi cuenta | Salir

Google docs Evaluación por parte de Guardado automáticamente el 2:10 p.m. GMT+02:00 Compartir ▾

Archivo Editar Ver Insertar Formato Formulario (12) Herramientas Ayuda

	A	B	C	D	E	F	G	H	I	J
	¿Cuál es su impresión global acerca de la tecnología Topic Map?	¿Cómo encontró la tarea de localizar revistas editadas por el CEPC con el Topic map?	¿Considera que los topic types seleccionados son adecuados para estructurar los contenidos de la biblioteca digital del CEPC?	Los topics empleados en el topic map para la biblioteca del CEPC, ¿son adecuados a sus demandas informativas?	Las occurrences empleadas en el topic map del CEPC, ¿se refieren a documentos adecuados de la colección?	Las associations empleadas en el topic map del CEPC, ¿son adecuados para la navegación entre las materias descriptivas de los contenidos de la colección?	Respecto al topic map de la NZETC, ¿qué le parece el topic map para el CEPC?	Respecto al topic map del Tax map, ¿qué le parece el topic map para el CEPC?	En cuanto a la visualización del topic map, ¿cuál prefiere de los tres topic maps?	Cree que la implantación del Topic map en una supuesta biblioteca digital semántica del CEPC mejora la calidad de sus servicios?
2	Fácil de utilizar, intuitivo y navegable. Útil, Apropiado para satisfacer mis demandas informativas	Difícil de buscar, Lento hasta dar con resultados, Ofrece resultados pertinentes a lo que se pide	De acuerdo	De acuerdo	De acuerdo	De acuerdo	Menos navegable e intuitivo, Más lento en la localización de información pertinente, Estructura peor los contenidos	Menos navegable e intuitivo, Más lento en la localización de información pertinente, Estructura peor los contenidos	Visualización del Topic map del Tax map	Si, muy de acuerdo
3	Fácil de utilizar, intuitivo y navegable. Útil, Apropiado para satisfacer mis demandas informativas	Rápido para localizar resultados, Ofrece resultados pertinentes a lo que se pide	Poco de acuerdo	Poco de acuerdo	De acuerdo	Poco de acuerdo	Menos navegable e intuitivo, Más lento en la localización de información pertinente, Estructura peor los contenidos	Menos navegable e intuitivo, Más lento en la localización de información pertinente, Estructura peor los contenidos	Visualización del Topic map del Tax map	Los topic maps aportan soluciones pero el construido no es adecuado
4	Fácil de utilizar, intuitivo y navegable. Útil	Difícil de buscar, Lento hasta dar con resultados	De acuerdo	Poco de acuerdo	Poco de acuerdo	Poco de acuerdo	Más navegable e intuitivo, Más rápido en la localización de información pertinente, Estructura mejor los contenidos	Menos navegable e intuitivo, Más lento en la localización de información pertinente, Estructura peor los contenidos	Visualización del Topic map del Tax map	Si, muy de acuerdo
	Fácil de utilizar						Más navegable e intuitivo. Más			

Añadir hoja Hoja 1 ▾ ¿Cuál es su impresión global

Listo Internet | Modo protegido: activado 100%

Figura 96. Cuestionario final que se envió a los usuarios con mayores habilidades. Elaboración propia.

El *topic map* finalmente diseñado para una supuesta biblioteca digital semántica jurídica del CEPC se muestra en la figura de a continuación.



Figura 97. Interfaz principal de la biblioteca digital semántica para el CEPC. Elaboración propia.

El resumen final de todos los elementos que constituyen el *topic map* se muestran en la siguiente pantalla tomada de la aplicación Ontopia con la que se realizó. Fundamentalmente es un *topic map* con un TAO de 48 *topic types* que agrupan 356 *topics*, 632 *associations* y 596 *occurrences*. Como vemos existen muchas más *associations* que *topics* siguiendo la idea de que para dotar de semántica a la navegación y localización de recursos informativos es importante definir varios tipos de relaciones a cada uno de los *topics*, así como determinar el papel que cada uno de estos van a ejercer en su relación con otros *topics*. Esto se expresa en la pantalla que se muestra a continuación como un resumen de la estructura de las *associations*.

Overall statistics		Association structure summary	
Topic Map Objects	#	Type	# Role types
Topics	356	Field in view	55 Fields view Field definition
Associations	632	Has association field	49 Role field Association field
Occurrences	596	Has association type	28 Association field Association type
Total TAOs	1584	Has cardinality	71 Cardinality Field definition
		Has datatype	26 Datatype Field definition
		Has field	192 Field owner Field definition
		Has identity type	3 Identity field Identity type
		Has name type	1 Name field Name type
		Has occurrence type	26 Occurrence type Occurrence field
		Has role type	49 Role field Association role type
		Is abstract	1 Topic type
		Is embedded view	5 Fields view
		Is hidden type	10 Ontology type
		Is hidden view	1 Fields view
		Organiza	1 Organizado por
		Superclass/subclass	40 Subclass Superclass
		Use View mode	6 Fields view View mode Field definition

Statistics for individual object types	
Topic Types	#
Number of different topic types	48
Actividades	2
Asociaciones científicas	1
Association field	28
Association role type	7
Association type	28
Becas	7
Biblioteca	2
Cardinality	4
Carta de servicios electrónicos	1
Convocatorias personal para el CEPC	1
Create action	3
Cursos extraordinarios	2
Datatype	7
Diploma	4
Documentación	3
Documentos Electrónicos	5
Edit mode	5
Fields view	8

Figura 98. Pantalla resumen de elementos del *topic map* para el CEPC. Elaboración propia.

La encuesta, orientada a conocer los aspectos considerados de mayor interés por parte de los usuarios, se realizó en el ámbito del CEPC con un universo de población de personas de ambos sexos, mayores de 18 años, licenciados y usuarios de los servicios de biblioteca del centro. Respecto al tamaño de la muestra, de los 18 usuarios para los que se diseñó la encuesta, sólo 12 han contestado finalmente a la misma acerca de la evaluación final del *topic map* como posible tecnología para la organización de una posible biblioteca digital semántica jurídica. Dado el bajo número de encuestados, no ha sido necesaria la realización de ningún tipo de ponderación ya que no tiene sentido una segmentación de la muestra. En cuanto a los puntos de muestreo, tomamos los mismos que para el primer cuestionario, las aulas y biblioteca del CEPC. El procedimiento de muestreo se ha efectuado mediante el envío a las personas que resolvieron las consultas sobre el diseño del *topic map* de los interfaces finales acerca de cuál fue el resultado final de este, pidiéndoles que contestaran a raíz de su experiencia cuáles eran los aspectos más destacados.

El error muestral⁸⁵⁹, al tratarse de un muestreo aleatorio simple, y teniendo en cuenta que la variable a estimar es una proporción, entonces el error de la muestra lo calculamos de acuerdo con la ecuación $e = 2 * \sqrt{p * q} \div n$. Como el error muestral que vamos a calcular

⁸⁵⁹ ALVIRA MARTÍN, F. *La encuesta: una perspectiva general metodológica*. Madrid : Centro de Investigaciones Sociológicas, 2004. pp.87-88.

lo queremos con un nivel de confianza del 95%, hemos multiplicado el algoritmo anterior por 2. Para nuestra encuesta vamos a considerar que $p=q$ (porcentajes de que sean o no una característica) e igual a 50% porque por igual es la probabilidad del resultado p y q. “n” es la muestra que para nuestro caso fueron únicamente 18 personas. Aceptando todas estas condiciones, diremos que el error muestral de la encuesta es el siguiente para un error muestral del 95%.

$$e=2 * \sqrt{50} * 50 \div 18 = \pm 23\%$$

Por tanto, obtenemos un muy elevado porcentaje de error para el conjunto del muestreo aleatorio simple debido al escaso número de la muestra. La fecha de realización de la misma fue entre el 02 y 23 de febrero de 2009.

6.5.2. Análisis de resultados de la encuesta.

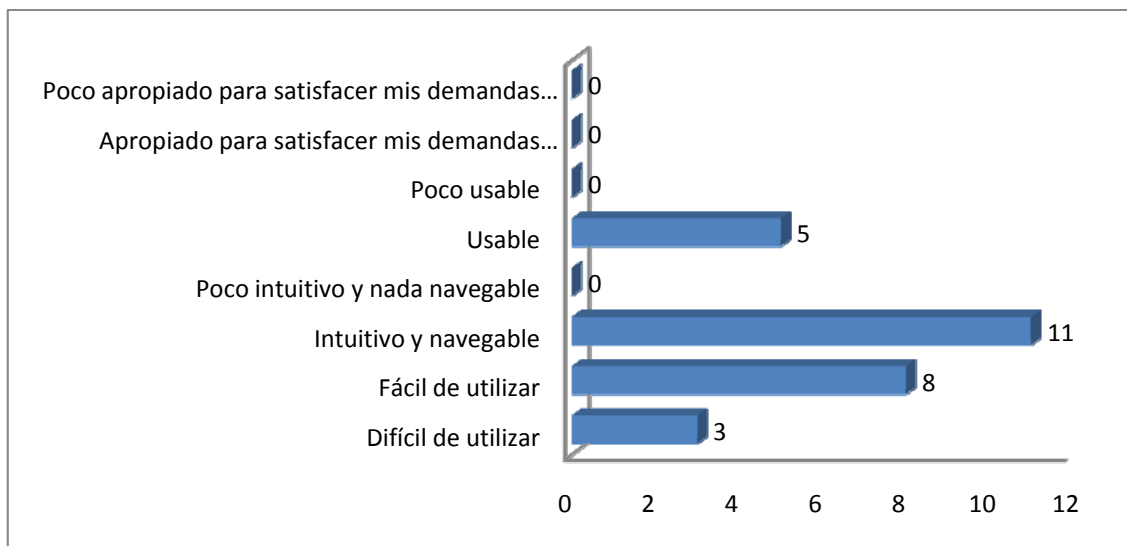
Hemos realizado 10 preguntas enviadas mediante correo electrónico que a su contestación devolvía los resultados a una base de datos para el análisis conjunto de las respuestas. Los resultados globales del cuestionario han sido los siguientes para cada una de las preguntas.

1. Respecto a la pregunta ¿Cuál es su impresión global acerca de la tecnología *Topic map*? La primera idea que obtenemos es que prácticamente para la totalidad de la muestra se trata de una tecnología intuitiva y navegable, así como fácil de usar. Son resultados muy optimistas que nos indican una amplia aceptación de esta tecnología por parte de los usuarios del CEPC, ya que ninguno de los encuestados indicó que pudiera tratarse de algo poco navegable, usable o inútil para la satisfacción de sus demandas informativas.

Difícil de utilizar	3	25%
Fácil de utilizar	8	67%
Intuitivo y navegable	11	92%
Poco intuitivo y nada navegable	0	0%
Usable	5	42%
Poco usable	0	0%
Apropiado para satisfacer mis demandas informativas	0	0%

Poco apropiado para satisfacer mis demandas informativas	0	0%
--	---	----

Tabla 21. Resultados de la pregunta acerca de la impresión global de la tecnología *Topic map*. Elaboración propia.

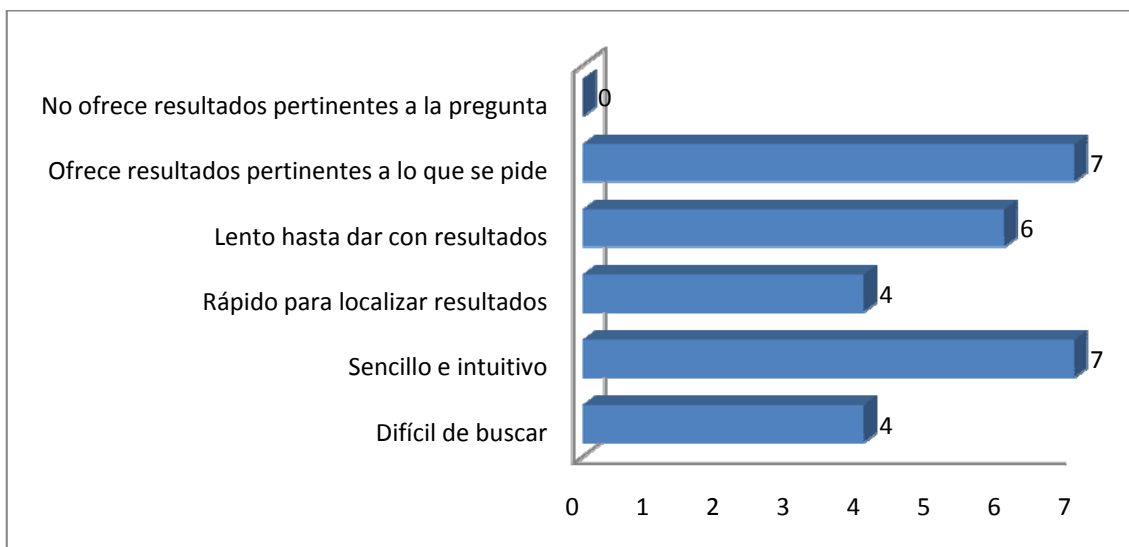


Gráfica 10. Sobre la impresión global de la tecnología *Topic map*. Elaboración propia.

2. Respecto a la pregunta ¿cómo encontró la tarea de localizar revistas editadas por el CEPC con el *topic map*? Esta segunda pregunta tenía por finalidad que los usuarios nos indicaran cómo perciben su capacidad para la búsqueda con *topic maps*. Usuarios que ya habían sido previamente seleccionados entre la muestra de usuarios con los que se trabajó en la mejora del *topic map*. De los resultados obtenemos dos ideas claras, y es que para ellos el empleo de esta tecnología permite la localización sencilla e intuitiva de información pertinente a las demandas informativas que realizaron. No obstante, se quejan de lentitud en el proceso de localización, seguramente porque las relaciones entre los *topics* todavía son escasas, lo que conlleva unas estrategias de navegación poco flexibles.

Difícil de buscar	4	33%
Sencillo e intuitivo	7	58%
Rápido para localizar resultados	4	33%
Lento hasta dar con resultados	6	50%
Ofrece resultados pertinentes a lo que se pide	7	58%
No ofrece resultados pertinentes a la pregunta	0	0%

Tabla 22. Respuestas de los usuarios a la localización de revistas del CEPC con *topic maps*. Elaboración propia.



Gráfica 11. Sobre la localización de revistas del CEPC con *topic maps*. Elaboración propia.

3. En cuanto a la pregunta relativa a si los *topic types* utilizados son adecuados para la estructurar la biblioteca digital semántica del CEPC, se realizó ya que como indicamos en esta tesis⁸⁶⁰, los *topic types* se corresponden con las clases de *topics* entorno a las cuales agruparemos todos los *topics* que añadamos durante el desarrollo del *topic map*. De las respuestas a esta pregunta tan sólo nos cabe afirmar que para nuestros usuarios están mayoritariamente de acuerdo con los que hemos empleado. Esto es lógico si pensamos que fundamentalmente la arquitectura de nuestro *topic map* la hemos diseñado análogamente a la arquitectura del sitio web del CEPC, ya que ésta fue diseñada conjuntamente tanto por la Subdirección General de Documentación y Publicaciones como por la Subdirección General de Estudios e Investigación, tal y como se indica en la carta de servicios electrónicos publicada por el CEPC⁸⁶¹.

Muy de acuerdo	1	8%
De acuerdo	9	75%
Poco de acuerdo	2	17%
No sabría qué responder	0	0%

⁸⁶⁰ Cfr.5.4.1.

⁸⁶¹ Carta de Servicios Electrónicos del Centro de Estudios Políticos y Constitucionales. En: http://www.cepc.es/include_mav/getfile.asp?IdFileImage=280. Consultado el 15/02/2010.

Tabla 23. Respuestas de los usuarios sobre si los *topic types* utilizados son adecuados para estructurar la biblioteca digital semántica del CEPC. Elaboración propia.

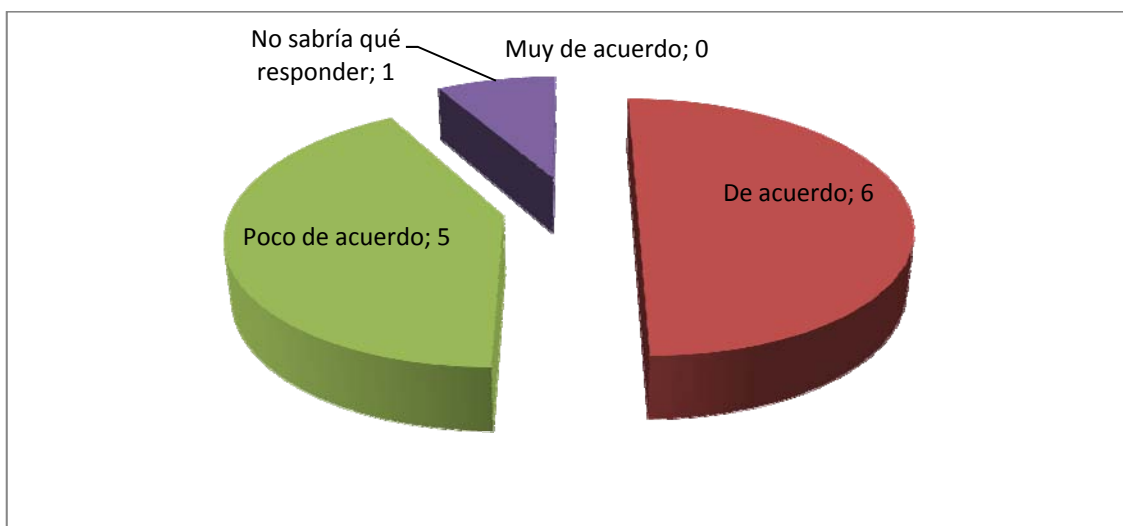


Gráfica 12. Sobre si los *topic types* utilizados son adecuados para estructurar la biblioteca digital semántica del CEPC. Elaboración propia.

4. No obstante, cuando realizamos la pregunta anterior referida a los *topics*, constatamos una falta de consenso al respecto. La idoneidad o no de los *topics* empleados para el *topic map* del CEPC es bastante discutida, si bien debemos tener en cuenta que los usuarios a quienes pasamos el cuestionario no pudieron realizar un empleo intenso del *topic map*. En general, la mitad de ellos estuvieron de acuerdo en los *topics* empleados y la otra mitad estuvo poco de acuerdo. Como comentamos, estos fueron seleccionados a partir del tesoro del ministerio de presidencia, así como de los distintos epígrafes de la arquitectura de sitio del CEPC. Sin embargo, tal y como nos comentó la profesora Gloria Gómez del Pulgar existen muchos usuarios que desconocen el empleo de tesauros para las búsquedas en catálogos y bases de datos lo cual nos hace replantearnos la idoneidad de emplear sus descriptores para construir un *topic map*.

Muy de acuerdo	0	0%
De acuerdo	6	50%
Poco de acuerdo	5	42%
No sabría qué responder	1	8%

Tabla 24. Respuestas de los usuarios respecto a la idoneidad de los *topics* empleados. Elaboración propia.

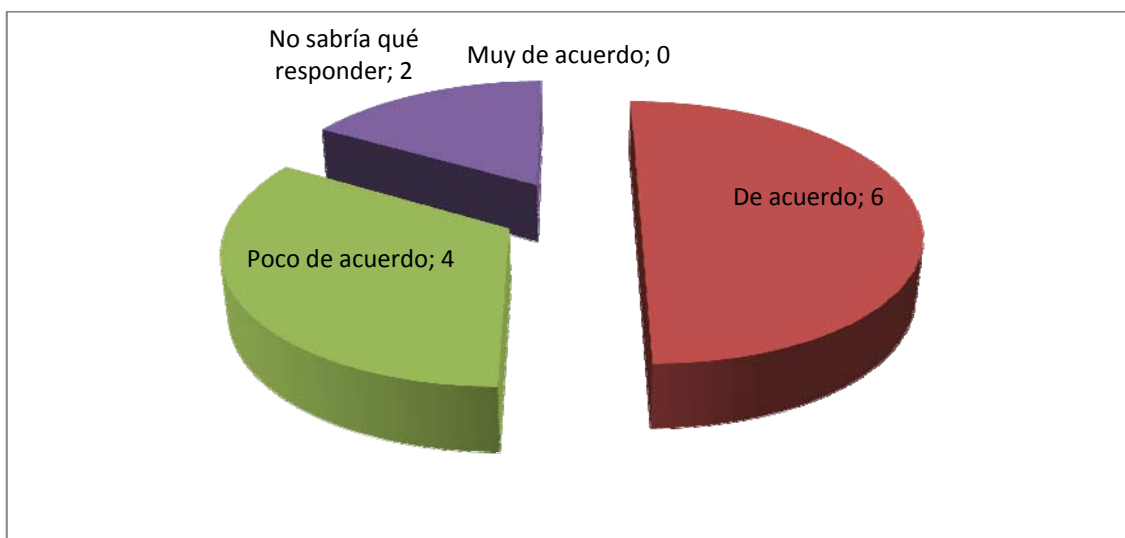


Gráfica 13. Sobre la idoneidad de los *topics* empleados. Elaboración propia.

5. Más evidente se hace la divergencia de impresiones de los usuarios cuando se les consulta acerca de la idoneidad de las *occurrences*, o enlaces a los documentos que mejor se adecúan a los *topics* seleccionados. La mitad de los mismos se muestran de acuerdo con los empleados y la otra mitad o está poco de acuerdo o no sabe qué responder, haciendo más evidente la falta de costumbre en el empleo del *topic map* que cuando se les preguntaba acerca de la idoneidad de los *topics*. A esto se añade la falta de conocimientos jurídicos del autor de esta tesis y del *topic map*, lo que nos debe hacer insistir en la necesaria labor interdisciplinar en el diseño de cualquier sistema documental jurídico, ya que un documentalista, aún en el caso de estar habituado al trabajo en una biblioteca jurídica, necesita del concurso de especialistas en los distintos campos del derecho que participan en la colección de una biblioteca digital jurídica.

Muy de acuerdo	0	0%
De acuerdo	6	50%
Poco de acuerdo	4	33%
No sabría qué responder	2	17%

Tabla 25. Respuestas de los usuarios a la idoneidad de las *occurrences* empleadas para los *topics* seleccionados. Elaboración propia.



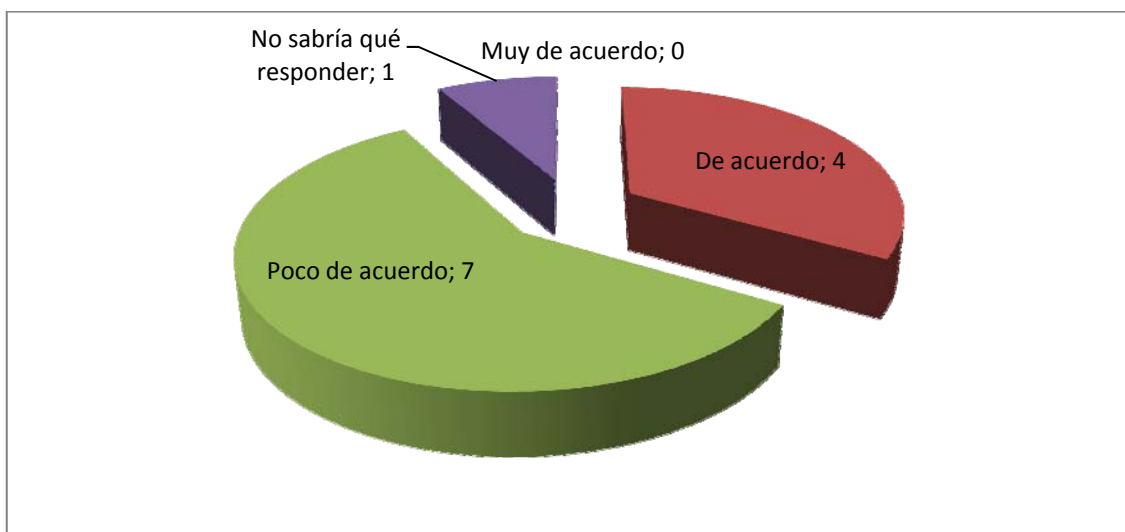
Gráfica 14. Acerca de la idoneidad de las *occurrences* empleadas para los *topics* seleccionados. Elaboración propia.

6. En lo relativo a la idoneidad de las *associations* empleadas para el *topic map* del CEPC que construimos para esta tesis, los usuarios nos indican en el cuestionario su disconformidad con las *associations* empleadas, seguramente este resultado tiene que ver con la lentitud que manifestaron en la realización de la prueba de localización de recursos informativos con el *topic map*. En efecto, el número de *associations* introducidas, 632, aunque alto, es insuficiente todavía para que el usuario pueda todavía localizar los recursos del CEPC de manera rápida e intuitiva. Esto quiere decir que para la constitución de una biblioteca digital semántica con unos pocos recursos informativos es esencial la declaración de un elevado número de *associations* para cada uno de los *topics*⁸⁶².

Muy de acuerdo	0	0%
De acuerdo	4	33%
Poco de acuerdo	7	58%
No sabría qué responder	1	8%

Tabla 26. Respuestas de los usuarios respecto de la idoneidad de las *associations* empleadas. Elaboración propia.

⁸⁶² Pepper, S. *The TAO of Topic Maps*. En: http://www.cepc.es/include_mav/getfile.asp?IdFileImage=280. Consultado el 15/02/2010.



Gráfica 15. Sobre la idoneidad de las associations empleadas según los usuarios. Elaboración propia.

Hasta ahora hemos visto preguntas realizadas a nuestros usuarios acerca del *topic map* para el CEPC en sí mismo. Ahora vamos a pasar a pedirles que nos indiquen cómo perciben nuestro *topic map* en relación con aquellos que les mostramos para indicarles de forma rápida de qué se trataba. El objetivo de estas preguntas ha sido hacer que los usuarios nos indiquen el grado de desarrollo de nuestro *topic map* en relación con otros que están más desarrollados y han contado con más personal y recursos para ello. Si bien es cierto que en ambos casos estructuran verdaderas bibliotecas digitales semánticas.

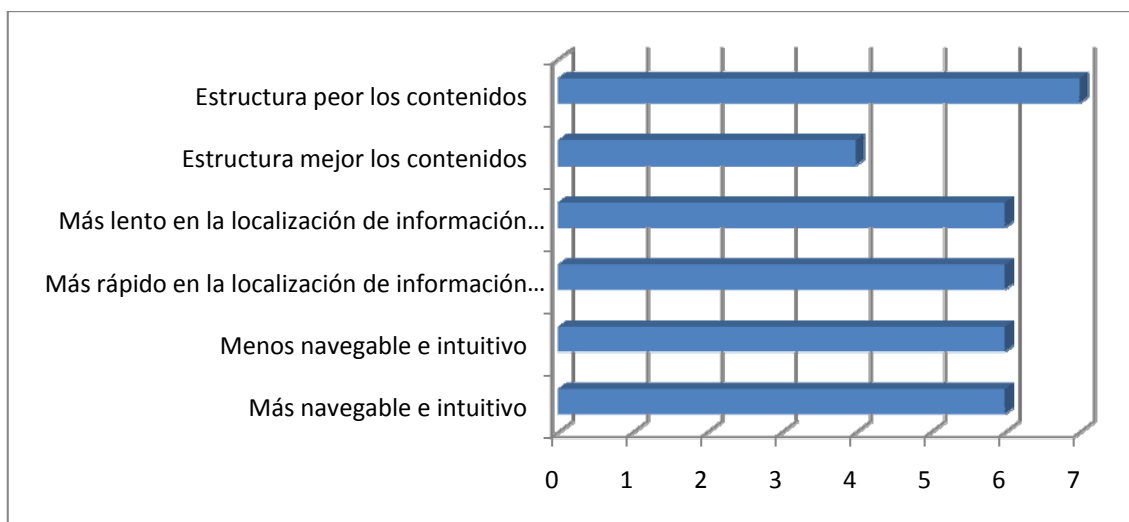
Asimismo, estos *topic maps* de referencia nos han servido para orientar a los usuarios acerca de la resolución de las consultas que se les pedía con el *topic map* en proceso de diseño. Estas preguntas son las siguientes.

7. Respecto a la comparación del *topic map* del CEPC con el *topic map* del New Zealand Text Centre constatamos una total diversidad de opiniones, salvo en que el *topic map* para el CEPC se le critica por tener los contenidos peor estructurados. La contestación al resto de preguntas refleja una situación muy incipiente de comprensión por parte de los usuarios de lo que es un *topic map*, así como una muestra de usuarios todavía insuficiente para poder extraer conclusiones más fiables.

Más navegable e intuitivo	6	50%
Menos navegable e intuitivo	6	50%

Más rápido en la localización de información pertinente	6	50%
Más lento en la localización de información pertinente	6	50%
Estructura mejor los contenidos	4	33%
Estructura peor los contenidos	7	58%

Tabla 27. Respuestas de los usuarios a la comparación entre el *topic map* del CEPC y el *topic map* del NZETC. Elaboración propia.



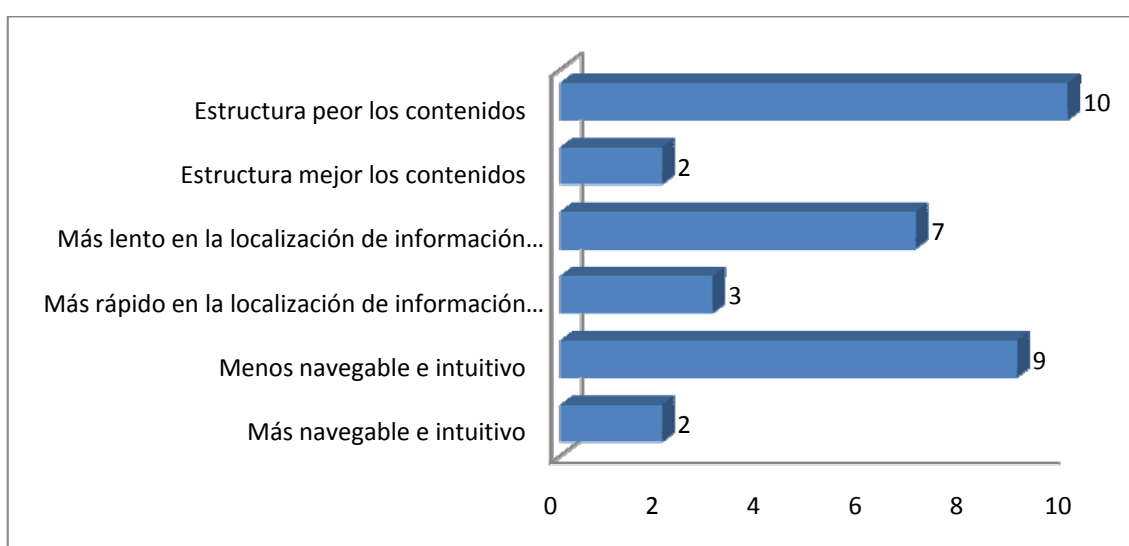
Gráfica 16. Acerca de la comparación realizada por los usuarios entre el *topic map* del CEPC y el *topic map* del NZETC. Elaboración propia.

8. En cuanto a la comparación del *topic map* del CEPC con el *topic map* del IRS Tax Map debemos decir que al tratarse esta de una aplicación con muchos años de sucesivas implementaciones en el servicio estadounidense de recaudación, goza de un nivel de desarrollo y evaluación muy superior a nuestro *topic map*. Esto hace que la gran mayoría de nuestros usuarios entiendan que el Tax Map es mucho más navegable e intuitivo que el nuestro, así como más rápido para la localización de información pertinente y con más capacidad para la estructuración de los contenidos. Hecho lógico si pensamos en las miles de *associations* con las que cuenta el Tax Map. Sin embargo, recogimos este caso porque no son muchas las publicaciones que constituyen su colección, únicamente son 95⁸⁶³.

⁸⁶³ BIEZUNSKI, M. *The internal revenue service tax map*. En: www.coolheads.com/MBPUBS/xmleurope2003/irstm4.ppt. Consultado el 15/02/2010.

Más navegable e intuitivo	2	17%
Menos navegable e intuitivo	9	75%
Más rápido en la localización de información pertinente	3	25%
Más lento en la localización de información pertinente	7	58%
Estructura mejor los contenidos	2	17%
Estructura peor los contenidos	10	83%

Tabla 28. Respuestas de los usuarios a la comparativa entre el *topic map* del CEPC y el IRS Tax Map. Elaboración propia.

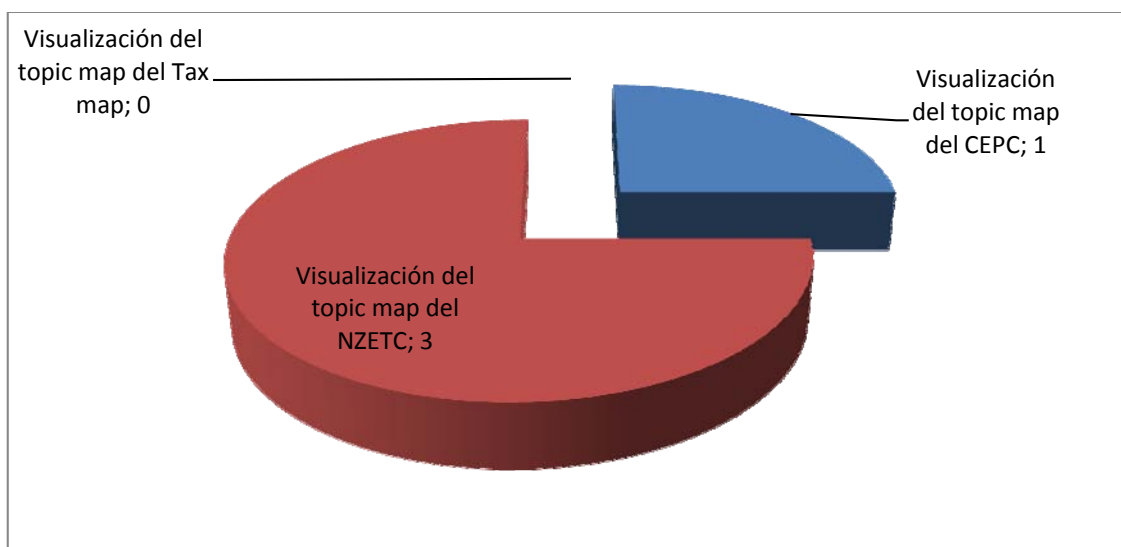


Gráfica 17. Sobre respuestas de los usuarios a la comparativa entre el *topic map* del CEPC y el IRS Tax Map. Elaboración propia.

9. Les hemos preguntado acerca del tipo de visualización que preferían entre los tres *topic maps*. Sin embargo, es este un aspecto demasiado técnico y de difícil comprensión para los mismos. Esto ha supuesto un nivel de respuesta muy poco significativo. No obstante, indicaremos cuáles han sido los resultados. De los doce usuarios que finalmente realizaron el cuestionario, tan sólo han respondido a esta pregunta 4 de ellos. Y de estos 4, tres de ellos han preferido la visualización del *topic map* para la biblioteca digital semántica de Nueva Zelanda. No podemos obtener conclusiones de estos resultados, pero si indicar que se trata de una visualización en índice pero que a diferencia del Tax Map se indican las relaciones semánticas entre los *topics*, lo que facilita la comprensión de los resultados y la navegación por parte del usuario.

Visualización del <i>topic map</i> del CEPC	1	8%
Visualización del <i>topic map</i> del NZETC	3	25%
Visualización del <i>topic map</i> del Tax map	0	0%

Tabla 29. Respuestas de los usuarios sobre visualización de los tres casos de *topic maps* mostrados. Elaboración propia.

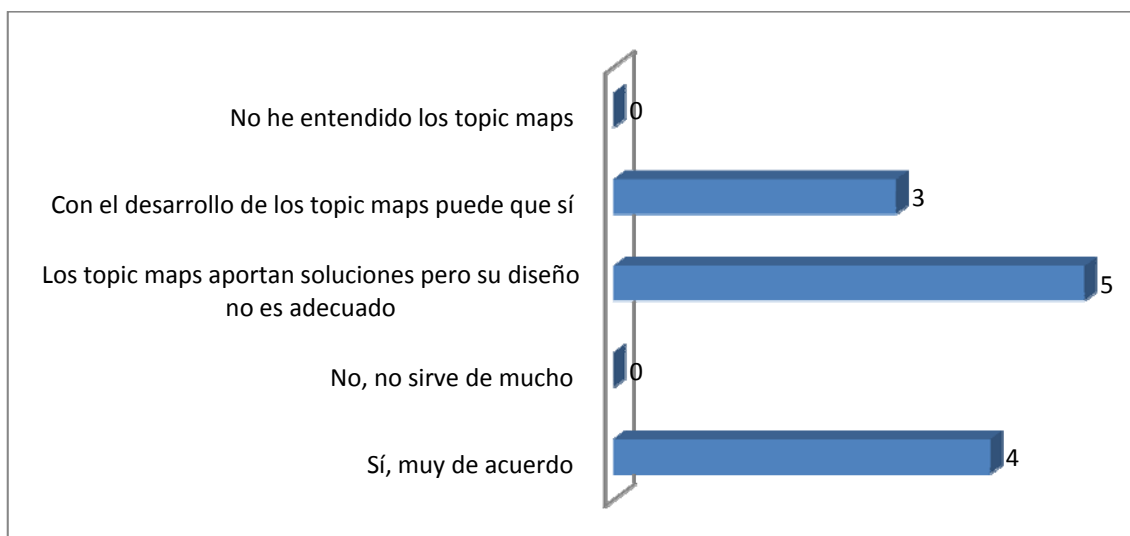


Gráfica 18. Sobre preferencias de visualización entre los tres casos de *topic maps* mostrados a los usuarios. Elaboración propia.

10. Finalmente, con el objeto de saber si nuestros usuarios estarían dispuestos a demandar la implantación en el centro de un *topic map* a medida que se vayan digitalizando los fondos de la biblioteca del CEPC, así como la implantación de servicios digitales en el marco de una biblioteca digital semántica. Estos, nos indican a la pregunta sobre si piensan que la implantación de un *topic map* puede mejorar los servicios de la biblioteca digital del CEPC, en su mayoría piensan que así es aunque critican el diseño que le hemos dado al *topic map*.

Sí, muy de acuerdo	4	33%
No, no sirve de mucho	0	0%
Los <i>topic maps</i> aportan soluciones pero su diseño no es adecuado	5	42%
Con el desarrollo de los <i>topic maps</i> puede que sí	3	25%
No he entendido los <i>topic maps</i>	0	0%

Tabla 30. Respuestas de los usuarios a la posibilidad de demandar la implantación de *topic maps* en la biblioteca digital jurídica del CEPC. Elaboración propia.



Gráfica 19. Acerca de la posibilidad de demandar la implantación de *topic maps* en la biblioteca digital jurídica del CEPC. Elaboración propia.

Capítulo 7

Conclusiones.

Globalmente, en la formulación de nuestras conclusiones, partimos de lo enunciado en nuestra hipótesis de partida: actualmente las bibliotecas digitales de dominios de conocimiento especializados, como el jurídico, carecen de la posibilidad de dotar a los usuarios de suficiente autonomía para poder satisfacer sus propias demandas de información. Y, sin embargo, esto sería hoy posible con la aplicación convergente de los servicios de la Web 2.0 y la Web 3.0. Si a esto añadimos la incipiente capacidad de los buscadores para la búsqueda por contexto, entenderemos la acuciante necesidad por adaptar el diseño de la arquitectura de información de las bibliotecas digitales a estas nuevas realidades, con las cuales nuestros usuarios ya están familiarizados particularmente pero no en el marco global de un sistema de información global como es una biblioteca digital.

Esperamos haber sido capaces de demostrar, fehacientemente, que las bibliotecas digitales semánticas, diseñadas con *Topic maps*, son un posible marco global para ofrecer conjuntamente los servicios de la Web 2.0 y Web 3.0 de forma coherente y adaptada adecuadamente al dominio de especialidad de la comunidad de usuarios a la que se desea dotar de autonomía informativa.

Sin embargo, hemos podido constatar durante la realización de esta tesis, que el desarrollo de las bibliotecas digitales no se está produciendo por igual ni en todos los campos del conocimiento ni en todo tipo de bibliotecas. Esto apunta a una nueva brecha digital en el futuro en cuanto al desarrollo de las bibliotecas digitales a causa del desarrollo más lento de la arquitectura de información en las mismas frente al rápido desarrollo de los servicios digitales en Internet. Es propósito de esta tesis haber sido capaces de reflejar las dificultades que encuentra la implantación de las tecnologías asociadas con la Web Semántica, y más concretamente con las ontologías y los *topic maps*, en la arquitectura de información de un nuevo modelo de biblioteca digital a través de la propuesta y evaluación por parte de sus potenciales usuarios de una biblioteca digital semántica jurídica; ya que, partimos de la confluencia de dos realidades presentes: i) las bibliotecas digitales especializadas con su lenguaje de especialidad y, ii) la implantación de un esquema específico de ontología, denominado *Topic Map*, para la estructuración de los contenidos en las bibliotecas digitales.

De la convergencia de ambas realidades, aplicadas a la biblioteca digital del Centro de Estudios Políticos y Constitucionales hemos llegado a las siguientes conclusiones teniendo en cuenta la hipótesis de la que partíamos.

1. **Pensábamos** que la organización del conocimiento en la biblioteca digital debía tener en cuenta tanto las características del documento como del dominio de conocimiento representado por su colección. En efecto, durante el diseño del mapa conceptual que nos sirvió de plantilla para realizar el *topic map* nos pareció imprescindible tener en cuenta, porque incluso el propio editor de *topic maps* así nos lo pedía, si los documentos que estábamos incluyendo eran revistas, libros o actas de congresos. Asimismo, el ámbito jurídico incluye una gran variedad de matices tanto en su propio campo como entre campos anexos como puedan ser la ciencia política o la historia contemporánea. Esto nos ha obligado necesariamente a contar con personas especializadas en cada uno de estos campos desde un primer momento. Estas personas no sólo nos han permitido diferenciar entre los distintos matices de los *topic types* designados para el *topic map* del CEPC, sino que tal y como explicamos en el capítulo tercero de la tesis podemos introducir términos de los distintos lenguajes de especialidad, entre los cuales es preponderante el lenguaje jurídico, que intervienen en la arquitectura de información de nuestra biblioteca digital semántica jurídica. Este lenguaje jurídico, tomado fundamentalmente de los tesauros de Derecho del Ministerio de Presidencia y de EUROVOC nos ha facilitado no sólo la labor de asignar adecuadamente los *topics* a los documentos, sino sobre todo decidir correctamente las *associations* entre los *topics*.
2. **Apostábamos**, y así lo hemos hecho, por un marco tecnológico-documental de *topic maps* que permite al usuario pasar a ser un protagonista activo en la resolución de sus demandas informativas y la construcción de su propio saber. En efecto, tal y como las encuestas del capítulo sexto nos han mostrado, los usuarios con la actual biblioteca digital del CEPC están restringidos a los modelos de búsqueda que se les impone a través de un catálogo y unos motores de búsqueda poco flexibles, especialmente cuando buscan documentos por materias. Además, la posibilidad que se les ofrece de utilizar un tesoro de Derecho para facilitarles las tareas de búsqueda no incrementa la satisfacción del usuario por desconocimiento del mismo acerca de su uso y por estar el tesoro separado de las herramientas de búsqueda.

En cambio, nuestros usuarios, a través de las entrevistas y de las encuestas acerca del empleo de nuestro *topic map* en la biblioteca digital semántica del CEPC nos muestran que el modelo flexible de búsqueda contextual por navegación es mucho más fácil de utilizar y les proporciona unos resultados más precisos y pertinentes. Y esto a pesar de las carencias en el diseño del *topic map* que nuestros usuarios nos han indicado porque cuando se les pedía realizar búsquedas en otros *topic maps* más desarrollados su satisfacción era mayor.

3. **Defendíamos** el empleo de un *topic map* a partir del cual construir una biblioteca digital semántica jurídica para el CEPC y así lo hemos hecho. De hecho, esta ha sido la materialización práctica de nuestra tesis. No se trataba de plantear un rediseño del sitio web para el CEPC, sino la creación de un marco ontológico global capaz de facilitar la navegación contextual entre los recursos informativos que constituyen la colección digital del CEPC, al mismo tiempo que se les facilita el acceso a la referencias bibliográficas de la colección en papel.
4. **Actuamos** a partir de la realización de un trabajo de campo cuanti-cualitativo, con entrevistas tanto al personal de biblioteca como a los usuarios, que junto a la realización entre la comunidad de usuarios de la biblioteca de dos encuestas enviadas a través de una aplicación de correo electrónico, Googledocs, nos ha permitido concluir que nuestra biblioteca digital semántica jurídica está destinada a satisfacer las demandas informativas de una comunidad de usuarios muy especializada pero con carencias notables en su alfabetización digital y en sus habilidades para la recuperación documental.
5. **Organizamos** la información obtenida del trabajo de campo por medio del empleo de un mapa conceptual. De la experiencia de su empleo hemos podido concluir que por sus características de modelo *node-arcs*, están especialmente adaptados para su conversión a un modelo de *Topic map*. Una vez concluido el mapa conceptual, utilizamos el editor de software libre de *topic maps* Ontopia para construir un *topic map* que bien puede ser utilizado como biblioteca digital semántica para el CEPC, con una visualización similar a la ofrecida en la organización de la información con el mapa conceptual.

6. **Erramos.** Centramos el *topic map* en el portal del CEPC por partir de una arquitectura de información ya diseñada para la web por los propios bibliotecarios y especialistas informáticos del centro para incluir de la manera más adecuada para su comunidad de usuarios todos los servicios de información que se ofrecen. Sin embargo, creemos que este pudo haber sido nuestro error en el diseño del *topic map*, puesto que los *topic maps* suponen un nuevo modelo de organización del conocimiento que poco tiene que ver con las arquitecturas de información conocidas hasta ahora para el diseño web; es decir, nuestro *topic map* ha arrastrado los defectos de la arquitectura web implantada actualmente en la biblioteca digital del CEPC y esto ha supuesto una clara limitación para lo que es el potencial de los *topic maps*. Quizás, tal y como propusimos en nuestra metodología se pueda construir en el futuro uno de ellos con todo el personal que creemos debería tomar parte en su puesta en marcha, de otra manera la realización individual de este trabajo colectivo lleva a un planteamiento quizás demasiado conservador para lo que es el potencial real de los *topic maps*.
7. **Recomendamos** el diseño de más *topic maps*, no sólo para bibliotecas digitales jurídicas necesariamente, si no para cualquier sistema de información en el campo del Derecho susceptible de hacerse interoperable con otros, al mismo tiempo que es capaz de reutilizar las incipientes ontologías jurídicas. El fin último que planteamos con esta tesis es el desarrollo de sistemas de información jurídica semánticos capaces de integrar progresivamente mas *topics* y *associations* para la navegación contextual multilingüe entre colecciones jurídicas, tal y como los profesionales del Derecho demandan en campos como el Derecho comparado.

Resumen.

La presente tesis se inserta en el marco de dos líneas de investigación recogidas en el plan estratégico de investigación del Departamento de Biblioteconomía y Documentación: 1) Análisis formal y de contenido, Lingüística y Sistemas de Conocimiento; 2) Tecnologías de la Información.

La hipótesis de partida ha sido que la documentación jurídica por sus características lingüísticas y documentales intrínsecas resulta adecuada para construir un *topic map* que dote

al usuario de autonomía para la satisfacción de sus demandas informativas en un contexto jurídico.

El objetivo ha sido valorar el potencial que tienen las bibliotecas digitales semánticas en las bibliotecas digitales jurídicas y cuáles son los problemas de su implantación. Para ello, con los recursos que la biblioteca digital del Centro de Estudios Políticos y Constitucionales pone a disposición de sus usuarios, hemos analizado las capacidades que estos tienen para utilizarlos con la idea de proponer un modelo, denominado *Topic map*, de recuperación por navegación contextualizada que les facilite adaptar el proceso de búsqueda a sus habilidades y necesidades en el marco de un nuevo modelo de biblioteca digital denominado biblioteca digital semántica jurídica.

La metodología empleada ha sido cuanti-cualitativa, si bien más cualitativa que cuantitativa. Se han realizado entrevistas personales tanto con el personal de la biblioteca como con sus usuarios. Asimismo, hemos recurrido a la realización de dos pequeñas encuestas entre los usuarios que incluían ejercicios de recuperación de información tanto con la actual biblioteca digital como con el *topic map* diseñado para la implantación de una biblioteca digital semántica jurídica. Todos los datos los hemos recopilado en un mapa conceptual que nos serviría como base para el posterior diseño del *topic map*.

Los resultados han sido fundamentalmente dos: a) una propuesta metodológica para la edición de *topic maps* para bibliotecas digitales semánticas jurídicas; b) hemos editado un *topic map* para el Centro de Estudios Políticos y Constitucionales que sirve como arquitectura de información para mejorar su actual diseño de sitio de acuerdo con las habilidades informacionales de sus usuarios. Tanto el autor de esta tesis como los propios usuarios hemos constatado la mejora significativa que introducen los *topic maps* en el empleo de los servicios de la biblioteca digital, a pesar de las carencias de desarrollo del *topic map* propuesto para el centro frente a otros del mismo contexto mostrados a los usuarios para su comparación. De esta comparación, hemos concluido que los *topic maps* constituyen un modelo ontológico con un enorme potencial para la recuperación semántica de información en el contexto de las ontologías jurídicas y las bibliotecas digitales semánticas jurídicas.

Summary of the Doctoral Thesis.

This thesis is within the framework of the two lines of scholarly investigation found in the strategic research plan of the Department of Library Science and Documentation: 1) Formal and content analysis, Linguistics and Knowledge Systems; 2) Information Technologies.

The starting hypothesis has been that legal documentation because of its intrinsic linguistic and documental features is appropriate for construction of a *topic map* which provides the users with the autonomy to meet informative demands within a legal context.

The aim is to assess the potential that semantic digital libraries have in legal digital libraries and determine what problems their implantation may pose. For that, with the resources that the digital library of the Center for Political and Constitutional Studies offers its users, we have analyzed the capabilities that they have to use them with the idea of proposing a model, termed, *Topic map*, for recuperation by contextualized navigation which facilitates adapting the search process to the abilities and needs within the framework of a new model of the digital library termed legal semantic digital library.

The methodology used is “quanti-qualitative”; perhaps more qualitative than quantitative. Personal interviews have been conducted with library personnel and library users. Likewise, we have employed short surveys with users that include information recuperation exercises with the current digital library as well as with the *topic map* designed for the implantation of a legal semantic digital library. We compiled all the data in a conceptual map which served as a basis for the design of the *topic map*.

There are two basic results: a) a methodology proposed for the edition of *topic maps* for legal semantic digital libraries; b) we have published a *topic map* for the Center of Political and Constitutional Studies that serves as information architecture to improve its current design of site according to the information abilities of the users. The author of this thesis and the users themselves have affirmed that there is a significant improvement brought about by its in the digital library service, despite the shortcomings in the development of the proposed *topic map* for the center in comparison with in a similar context, shown to the users for the purposes of comparison. From this comparison, we have concluded that the

topic maps make up an ontological model with enormous potential for semantic information recuperation within the context of legal ontologies and legal semantic digital libraries.

Conclusions.

Globally, in the formulation of our conclusions, we leave of that enunciated in our departure hypothesis: currently the digital libraries of specialized domains of knowledge, as the juristic one, lack the possibility of endowing to the users of enough autonomy to be able to satisfy their own lawsuits of information. And, however, this would be possible today with the convergent application of the services of the Web 2.0 and the Web 3.0. If to this we add the incipient capacity of the searchers for the search for context, we will understand the pressing necessity to accommodate the design of the architecture of information from the digital libraries to these new realities, with which our users are already familiarized particularly but not in the global frame of a global information system as it is a digital library.

We wait to have been able to demonstrate, convincingly that the digital semantic libraries, designed with Topic maps, are a global possible frame to offer the services of the Web jointly 2.0 and Web 3.0 in a consistent and adapted way appropriately to the domain of specialty of the user community to the one that wants you to endow of information autonomy.

However, we have been able to verify during the realization of this thesis that the development of the digital libraries is not taking place equally neither in all the fields of the knowledge neither in all type of libraries. This points to a digital new breach in the future as for the development of the digital libraries because of the most dilatory development in the architecture of information in the same ones in front of the express development of the digital services in Internet. It is purpose of this thesis to have been able to reflect the difficulties that it finds the installation of the technologies associated with the Web Semantics, and more concretely with the ontologies and the topic maps, in the architecture of information of a new model of digital library through the proposal and evaluation on the part of their potential users of a library artificial digital semantics; since, we leave of the fork of two present realities: i) the digital specialized libraries with their language of specialty and, ii) the installation of a specific outline of ontology, denominated Topic Map,

for the structuring of the contents in the digital libraries. Of the convergence of both realities, applied to the digital library of the Center of Political and Constitutional Studios we have reached the following conclusions keeping in mind the hypothesis of which we left.

1. We **thought** that the organization of the knowledge in the digital library should keep so much in mind the features of the document like of the domain of represented knowledge for its collection. Indeed, during the design of the conceptual map that served us as template to carry out the topic map we found indispensable to keep in mind, because the own editor of topic maps even requested this way it to us, if the documents that we were including were magazines, books or records of congresses. Also, the juristic environment includes a wide variety of hues so much in its own field as among annexed fields as they can be the political science or the contemporary history. This has necessarily put under an obligation to have people specialized in each one of these fields from a first moment. These people have not only allowed us to differ among the different hues of the topic types designated for the topic map of CEPC, but rather such and like we explain in the chapter third of the thesis we can insert terms of the different languages of specialty, among which it is preponderant the juristic language that you intervene in the architecture of information of our library artificial digital semantics. This juristic language, taken fundamentally of the thesauruses of Right of the Ministry of Presidency and of EUROVOC it has not only facilitated us the work of allocating the topics appropriately to the documents, but mainly to decide the associations correctly among the topics.
2. We **bet**, and we have made this way it, for a technological-documental frame of topic maps that allows to the user to become an enabled main character in the resolution of their information lawsuits and the construction of their own knowledge. Indeed, such and as the surveys of the chapter sixth have shown us, the users with the digital current library of CEPC are restricted to the search models that are imposed through a catalog and some not very floppy search engines, especially when they search documents for matters. Also, the possibility that you/they are offered of using a thesaurus of Right to facilitate

them the search tasks doesn't increase the user's satisfaction for ignorance of the same one about its use and the thesaurus separated from the search tools to be. On the other hand, our users, through the interviews and of the surveys about the employment of our topic map in the library digital semantics of CEPC show us that the floppy pattern of contextual search for sailing is much easier of using and it provides them some more precise and more pertinent results. And this in spite of the lacks in the design of the topic map that our users have indicated us because when they were asked to carry out searches in other topic more developed maps their satisfaction it was bigger.

3. We **defended** the employment of a topic map starting from the one which to build a library artificial digital semantics for CEPC and we have made this way it. In fact, this it has been the practical materialization of our thesis. It was not about outlining a redesign of the website for CEPC, but the creation of an ontologic global frame able to facilitate the contextual sailing at the same time among the information petitions that constitute the digital collection of CEPC, that they are facilitated the access to the bibliographical references of the collection in paper.
4. We **proceed** starting from the realization of a quanti-qualitative fieldwork, with interviews so much to the library personnel like the users that next to the realization among the user community of the library of two surveys correspondents through an electronic mail application, Googledocs, it has allowed us to conclude that our library artificial digital semantics is dedicated to satisfy the information lawsuits of a very specialized user community but with remarkable lacks in its digital literacy and in its abilities for the documental retrieval.
5. We **organize** the obtained information of the fieldwork by means of the employment of a conceptual map. Of the experience of their employment we have been able to conclude that for their features of model node-arcs, they are specially adapted for their conversion to a model of Topic map. Once concluded the conceptual map, we use the software publisher free of topic maps Ontopia to build a topic map that well can be used as library digital

semantics for CEPC, with a similar visual display to the one offered in the organization of the information with the conceptual map.

6. We **mistake**. We center the topic map in the portal of CEPC to already leave of an architecture of information designed for the web by the own librarians and computer experts of the center to include in the most appropriate way for their user community all the information services that offer. However, we believe that this it could have been our error in the design of the topic map, since the topic maps supposes a new model of organization of the knowledge that little it has to do up to now with the well-known architectures of information for the design web; that is to say, our topic map has dragged the bugs of the architecture web implanted currently in the digital library of CEPC and this has supposed a clear limitation for what is the potential of the topic maps. Maybe, such and like we offered in our methodology you can build in the future one of them with the whole personnel that we believe it should participate in their setting in march, otherwise the individual realization of this joint work takes maybe to a position too conservative for what is the actual potential of the topic maps.
7. We **recommend** the design of more topic maps, it doesn't only stop necessarily digital juristic libraries, if it doesn't stop any information system in the field of the susceptible Right of becoming interoperable with other, at the same time that it is capable of reusing the juristic incipient ontologies. The end last that we outline with this thesis are the artificial semantic development of information systems able to integrate progressively but topics and associations for the sailing contextual multilingual among juristic, such collections and as the professionals of the Right they sue in fields like the Comparative jurisprudence.

Bibliografía

A

AARTS, M. *Knowledge models for government information*. En: <http://www.topicmaps.com/tm2008/aarts.pdf>. Consultado el 15/07/2009.

AGNEW, G.: Part 2. How libraries can employ metadata. *Cataloging & Classification Quarterly*, vol.36, nº3/4, 2003.

AGOSTI, M. [et al.]. *Evaluation and Comparison of the service architecture , P2P, and grid approaches for DLs*. Munich : DELOS, 2006. En: www.delos.info Consultado el 14/12/2009.

AHMED, K. *Topic maps. A practical introduction with case studies*. En: https://ssl.bnt.com/idealliance/papers/xmle02/dx_xmle02/papers/03-05-01/03-05-01.pdf. Consultado el 15/12/2009.

AHMED, K. Topic maps for repositories. En *XML Europe Proceedings*, 2000. En: <http://www.gca.org/papers/xmleurope2000/pdf/s29-04.pdf>. Consultado el 15/07/2009.

AHMED, K. [et al.]. *Professional XML Metadata*. Chicago : Wrox Press, 2001.

AHMED, K. Beyond PSIs. Topic map design patterns. En *Extreme Markup*, 2003. En: www.mulberrytech.com/Extreme/Proceedings/html/2003/Ahmed01/EML2003Ahmed01.html. Consultado el 15/12/2009.

AHMED, K. TMSHare. Topic map Fragment Exchange in a Peer to Peer application. En *Proceedings of XML Europe 2003*. En: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.922&rep=rep1&type=pdf>. Consultado el 15/12/2009.

AHMED, K. *Topic map patterns for information architecture*. En: www.techquila.com/tmsinia.html. Consultado el 15/02/2010.

AHMED, K. ; MOORE, G. An introduction to Topic Maps. *Microsoft Architect Journal*, nº5, 2005. En: <http://msdn.microsoft.com/en-us/library/aa480048.aspx>. Consultado el 15/02/2010.

ATTCHISON, J. *Thesaurus Construction and use: a practical manual*. London: Aslib, 1990.

ATTCHISON, J. ; BAWDEN, D. ; GILCHRIST, A. *Thesaurus Construction and Use: A Practical Manual*. Chicago/London: Fitzroy Dearborn, 1971.

ATTCHINSON, J. ; GILCHRIST, A. *Thesaurus construction and use: a practical manual*. Londres : Aslib, 1990.

ATTCHISON, J. ; GILCHRIST, A. ; BAWDEN, D. *Thesaurus construction and use: a practical manual*. London : Europa Publications, 2000.

ALCARAZ VARÓ, E. ; HUGHES, B. *El Español Jurídico*. Barcelona : Ariel, 2002.

ALGERMISSEN, J. *Gooseworks.org*. En: www.gooseworks.org/index.html. Consultado el 14/12/2009.

ALGERMEISSEN , J. *Mapalizer*. En: www.topic_mapping.com/mapalizer. Consultado el 15/12/2009.

ALGERMISSEN, J. *V topic map browser*. En: http://www.topic_mapping.com/v.html. Consultado el 12/10/2009.

ALTHEIM, M. *Ceryle*. En: <http://www.altheim.com/ceryle/>. Consultado el 15/12/2009.

ALVES REZENDE, E. *Mapa de Tópicos-MT*. Director: Maria Luiza Machado Campos. [Tesis no publicada]. Universidade Federal do Rio de Janeiro, Núcleo de Computacao Eletrônica, 2003.

ALVITE DÍEZ, M^a. Las herramientas terminológicas en los sistemas de información jurídica. *Scire*, vol.10, nº1, 2004.

ALVIRA MARTÍN, F. *La encuesta: una perspectiva general metodológica*. Madrid : Centro de Investigaciones Sociológicas, 2004.

ALVITE DÍEZ, M^a. Tendencias de la investigación sobre recuperación de información jurídica. *Revista Española de Documentación Científica*, vol.26, nº2, 2003.

AMAT I NOGUERA, N. *La biblioteca electrónica*. Madrid : Fundación Germán Sánchez Ruipérez, 1990.

ANSI/NISO Z39.19-2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. En: <http://webs.um.es/iscil/Z39-19-2005.pdf>. Consultado el 15/12/2009.

ARNTZ, R. ; PICT, H. *Introducción a la terminología*. Madrid : Fundación Germán Sánchez Ruipérez, 1995.

ASARO, C. ; BIASIOTTI, M. A. ; GUIDOTTI, P. ; PAPINI, M. ; SAGRI, M. T. ; TISCORNIA, D. A domain ontology : Italian crime ontology. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies & Web based legal information management*, 2003. En: <http://www.lri.jur.uva.nl/~winkels/LegOnt2003/Asaro.pdf>. Consultado el 12/12/2009.

AUSTIN, D. *PRECIS: a manual of concept analysis and subject indexing*. London : Council of the British National Bibliography, 1974.

AZOFRA MÁRQUEZ, M^a J. *Cuestionarios*. Madrid : CIS, 1999.

B

BACA, M. Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information. *Cataloging & Classification Quarterly*, vol.36, nº3/4, 2003. pp. 47-55.

BAKER, Collin F. ; FILLMORE, Charles J. ; LOWE, John B. The Berkeley FrameNet project. En *Proceedings of the COLING-ACL*, Montreal, Canada. 1998.

BANDHOLTZ, T. A Taxi in Knowledge Land: A Use Case that Combines Topic maps and Web Services in a Public Portal. En *Proceedings of Extreme Markup Languages 2002*. En: <http://www.bandholtz.info/publications/2002-XMLE-TB.pdf>. Consultado el 15/12/2009.

BARDIN, L. *El análisis de contenido*. Madrid : Akal, 1986.

BARTA, R. *AsTMa*. En: <http://astma.it.bond.edu.au/astma=-spec-xtm-1.0r1.8.dbk>. Consultado el 15/12/2009.

BARTA, R. *TMIP, a RESTful Topic maps Interaction Protocol*. En: www.mulberrytech.com/Extreme/Proceedings/html/2005/Barta01/EML2005Barta01.html. Consultado el 15/12/2009.

BARTA, R. *Is he the one? Subject Identification in Topic maps*. En: <http://topicmaps.it.bond.edu.au/docs/21>. Consultado el 15/12/2009.

BATER, B. Topic Maps: Indexing in 3-D. En GILCHRIST, A. ; MAHON, B. *Information architecture: designing information environments for purpose*. New York : Neal-Schuman Publishers, 2004.

BATLEY, S. *Information Architecture for Information Professionals*. Oxford : Chandos Publishing, 2007.

BELKIN, N. J. ; MARCHETTI, P. BRAQUE: design of an interface to support user interaction in information retrieval. *Information Processing and Management*, vol. 29, nº3, 1993.

BENGOECHEA, L. ; PATRICIO, M. A. Sistemas de visualización para bibliotecas digitales. *Revista Española de Documentación Científica*, vol. 28, nº3, 2005.

BENJAMINS, V. R. Law and the Semantic Web, an Introduction. En *Law and the semantic web: legal ontologies, methodologies, legal information retrieval, and applications*. Berlin : Springer, 2005.

BENJAMINS, V.R. (et al.). *Luriservice: Un FAQ inteligente para los jueces en su primer destino*. En: <http://www.sepln.org/revistaSEPLN/revista/31/31-Pag347.pdf>. Consultado el 15/12/2009.

BENJAMINS, V.R. ; CASANOVAS, P. ; BREUKER, J. ; GANGEMI, A. *Law and the semantic web, an introduction*. Berlin : Springer-Verlag, 2005.

BENJAMINS, V. R. ; CONTRERAS, J. ; CASANOVAS, P. ; AYUSO, M ; BECUE, M ; LEMUS, L. ; URIOS, C. Ontologies of profesional legal knowledge as the basis for intelligent IT support for judges. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies&Web based legal information management*, 2003.

En: www.lri.jur.uva.nl/~winkels/LegOnt2003/Benjamins.pdf. Consultado el 12/12/2009.

BENJAMINS, R.; CASANOVAS, P.; CONTRERAS, J.; LÓPEZ COBO, J. M.; LEMUS, L. Iuriservice: An Intelligent Frequently Asked Questions System to Assist Newly Appointed Judges. En BENJAMINS, V.R., CASANOVAS, P. BREUKER, J., GANGEMI, A. (Eds.). *Law and the Semantic Web. Legal Ontologies. Methodologies, Legal Information Retrieval, and Applications*. Berlin : Springer Verlag, 2005, pp. 201-217.

BERNERS-LEE, T. Tejiendo la red: el inventor del World Wide Web nos descubre su origen. Madrid : Siglo Veintiuno, 2000.

BERNERS-LEE, T. ; HENDLER, J. ; LASSILA, O. The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American Magazine*, 2001. En: <http://www.sciam.com/article.cfm?id=the-semantic-web&print=true>. Consultado el 15/02/2010.

BERNERS-LEE,T ; HENDLER, J. ; LASSILA, O. The Semantic Web. *Scientific American*, vol.284, n°5, 2001. En: <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>. Consultado el 15/02/2010.

BIEZUNSKI, M. The story of a Topic maps Use Case: The IRS Call Center Tax Map. En *Proceedings of XML Europe 2003*. London: Idealliance, 2003. En: www.coolheads.com/MBPUBS/xmleurope2003/irstm4.ppt. Consultado el 15/12/2009.

BIEZUNSKI, M. *A matter of perspectives: talking about topic maps*. En: www.mulberrytech.com/Extreme/Proceedings/html/2005/Biezunski01/EML2005Biezunski01.html. Consultado el 15/12/2009.

- BIEZUNSKI, M. *The internal revenue service Tax Map*. En: www.coolheads.com/MBPUBS/xmleurope2003/irstm4.ppt. Consultado el 15/02/2010.
- BIEZUNSKI, M. The gap between structured and unstructures needs a bridge. En *Extreme Markup Conference*, 2002. En: <http://www.coolheads.com/MBPUBS/extreme02/index.htm> Consultado el 15/12/2009.
- BIEZUNSKI, M. Introduction to the Topic Maps Paradigm. En PARK, J. *XML Topic Maps. Creating and using Topic Maps for the Web*. Boston : Addison-Wesley, 2002.
- BIEZUNSKI, M. ; NEWCOMB, S. *RDF and Topic Maps: Something new for everyone* En: <http://conferences.idealliance.org/extreme/html/2001/Biezunski01/EMI2001Biezunski01.html>. Consultado el 15/12/2009.
- BIEZUNSKI, M. ; NEWCOMB, S. ; BRYAN, M. *Guide to the topic map standardization*. En: <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0323.htm>. Consultado el 15/12/2009.
- BING, J. *Knowledge-based Systems for Representing Legal Norms*. En: http://www.topicmaps.com/tmc/presentation.jsp?conf=TM2009&id=Jon_Bing~TM2009~Knowledge-based_Systems_for_Representing_Legal_Norms. Consultado el 15/07/2009.
- BLYBERG, J. *Library 2.0 websites: Where to begin?* En: www.blyberg.net Consultado el 14/12/2009.
- BLOG ELECTRICFOREST. En: <http://www.altheim.com/ef>. Consultado el 15/12/2009.
- BOER, A. ; HOEKSTRA, R. ; WINKELS, R. The CLIME ontology. En *Proceedings of the second international workshop on legal ontologies*. Amsterdam, 2001. En: <http://www.lri.jur.uva.nl/jurix2001/papers/Boer.pdf> Consultado el 12/12/2009.
- BOGDAN, R. Foreword. En MELLON, C.A. *Naturalistic inquiry for Library science: methods and applications for research, evaluation, and teaching*. Nueva York : Greenwood Press, 1990.

BOND UNIVERSITY. En: <http://topic.maps.bond.edu.au>. Consultado el 15/12/2009.

BORGMAN, C. L. *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*. Cambridge : MIT Press, 2000.

BONURA, L. S. *The art of indexing*. New York : Wiley, 1994.

BRADLEY, J. Methodological issues and practices in qualitative research. *Library Quarterly*, vol.63, nº4, 1993.

BRADLEY, J ; SUTTON, B. Qualitative research. *Library Quarterly*, vol.63, nº4, 1993.

BRADY, T. Representing Software System Information in a Topic map. En *Extreme Markup Conference*, 2004. En: www.mulberrytech.com/Extreme/Proceedings/html/2004/Brady01/EML2004Brady01.html. Consultado el 15/12/2009.

BREUKER, J. ; VALENTE, J. ; WINKELS, R. Use and reuse of legal ontologies in Knowledge Engineering and Information Management. En *Law and the Semantic Web*. Berlin : Springer, 2005.

BROPHY, P. *The library in the twenty-first century*. London : Facet, 2005.

BROUGHTON, V. *Essential thesaurus construction*. London : Facet Publishing, 2006.

BROWN, D. ; BIEZUNSKI, M. *IRS Tax Map. Electronic research tool for tax law assistants*. En: http://ipgems.com/present/06_Brown-Biezunski_10-03.pdf. Consultado el 15/02/2010.

BROWN, P. *Information Architecture with XML. A management strategy*. Chichester : Wiley, 2003.

BUBL. En: www.bubl.ac.uk. Consultado el 01/02/2010.

BURGO Y MARCHÁN, Ángel Martín del. *El lenguaje del derecho*. Barcelona: Bosch, 2000.

BUSH, V. As we may think. *Atlantic Monthly*, nº176, 1945, pp.101-108. En:
<http://www.theatlantic.com/doc/194507/bush>. Consultado el 15/02/2010.

C

CABRÉ CASTELLVÍ, M.T. *La terminología. Teoría, metodología, aplicaciones*. Barcelona : Ediciones Antártida, 1993.

CABRÉ CASTELLVÍ, M.T. *La terminología. Representación y comunicación. Una teoría de base comunicativa y otros artículos*. Barcelona : Institut Universitari de Lingüística Aplicada. Barcelona : Universitat Pompeu Fabra, 1999.

CAMPOS, M.L.A. *Linguagem documentária: teorias que fundamentam sua elaboracao*. Niteroi : Eduff, 2001.

CAMPOS PARDILLOS, M.A. El lenguaje de las ciencias jurídicas: nuevos retos y nuevas visiones. En Alcaraz Varó, E. ; Mateo Martínez, J. ; Yus Ramos, F. *Las lenguas profesionales y académicas*. Barcelona : Ariel, 2007.

CAPLAN, P. *Metadata Fundamentals for All Librarians*. Chicago : American Library Association, 2003.

CARTA DE SERVICIOS ELECTRÓNICOS. En:
http://www.cepc.es/include_mav/getfile.asp?IdFileImage=280. Consultado el 15/02/2010.

CASANOVAS, P. *Ontologías jurídicas profesionales. Sobre “conocer” y “representar” el derecho*. En:
http://www.leibnizsociedad.org/secciones/mater/pon/textos/ontologias_pompeu.pdf. Consultado el 15/12/2009.

CASANOVAS, P. [et al.] *Semantic Web: a legal case study*. En: Davies, J. ; Studer, R. ; Warren, P. *Semantic Web Technologies*. Chichester : Wiley, 2006.

CASTELLS, M. *La sociedad red*. Madrid : Alianza, 2005.

CENTRO DE ESTUDIOS POLÍTICOS Y CONSTITUCIONALES. CEPC. En: <http://www.cepc.es>. Consultado el 15/12/2009.

CHAUMIER, J. *Les techniques documentaires*. Paris : PUF, 1974.

CHOMSKY, N. *El lenguaje y el entendimiento*. Barcelona: Planeta-Agostini, 1980.

CHOWDHURY, G. G ; CHOWDHURY, S. *Introduction to Digital Libraries*. London: Facet Publishing, 2003.

CHODOROW, S. ; LYMAN, P. The responsibilities of Universities. En HAWKINS, B. L. ; BATTIN, P. *The mirage of continuity: reconfiguring academic information resources for the 21st century*. Boston : Council on Library and Information Resources and Association of American Universities, 1998.

CIANCARINI, P. ; GENTILUCCI, R. ; PIRRUCCIO, M. ; PRESUTTI, V. ; VITALI, F. Metadata on the web: On the integration of RDF and Topic Maps. En *Proceedings of the Extreme Markup Languages 2003*. En: <http://conferences.idealliance.org/extreme/html/2003/Presutti01/EML2003Presutti01.html>. Consultado el 14/12/2009.

CLEVELAND, G. Digital libraries: Definitions, issues and challenges. *IFLANET: Activities & Services*. En: www.ifla.org/VI/5/op/udtop8/udtop8.htm. Consultado el 15/12/2009.

CODINA, LL. ; PEDRAZA-JIMÉNEZ, R. ; ROVIRA, C. Ontologías y sistemas de información documental. En CODINA, LL. ; MARCOS, M.C. ; PEDRAZA, R. *Web semántica y sistemas de información documental*. Gijón : Trea, 2009

CONCLIN, J. Hypertext: an introduction and survey. *IEEE Computer*, vol. 20, nº9, septiembre 1987.

CONTRERAS, J. (et al.). A semantic portal for the international affairs sector. En *Proceedings of EKAW 2004: engineering knowledge in the age of the semantic web*. Berlin : Springer, 2004.

COOLHEADS CONSULTING. En: <http://www.coolheads.com/publications.htm>.

Consultado el 15/12/2009.

COULLING, K. *Quality management for information and library managers*. London : Gower, 1996.

COVE, J. F. ; WALSH, B. C. Online text retrieval via browsing. *Information Processing and Management*, vol. 24, nº1, 1988.

CRANE, G. What do you do with a million books? *D-Lib magazine*, vol.12, nº3. En: www.dlib.org/dlib/march06/crane/03crane.html. Consultado el 15/02/2010

CREGAN, A. *Building Topic maps in OWL-DL*. En: www.mulberrytech.com/Extreme/Proceedings/html/2005/Cregan01/EML2005Cregan01.html. Consultado el 15/12/2009.

CUETO APARICIO, Marina. El nuevo tesoro EUROVOC como instrumento de recuperación multilingüe. En *II Jornadas de Tratamiento y Recuperación de la Información*. Leganés : Universidad Carlos III, 2003.

CURRÁS, E. *Tesauros: manual de construcción y uso*. Madrid : Kaher II, 1998.

CURRÁS, E. *Ontologías, taxonomías y tesauros*. Gijón : Trea, 2005.

D

DACONTA, M. ; OBRST, L. ; SMITH, K. *The Semantic Web. A guide to the future of XML, Web Services, and Knowledge Management*. Indianapolis : Wiley Publishing, 2003.

DATA MODEL. En: www.isotopic maps.org/sam/sam-model. Consultado el 15/10/2009.

DECKER, S. ; LACHER, M. *On the Integration of Topic Maps and RDF Data*. En: www.semanticweb.org/SWWS/program/full/paper53.pdf. Consultado el 14/12/2009.

DECRETO de 9 de septiembre de 1939. Artº 1º. BOE de 11 de septiembre de 1939. En: <http://www.boe.es/datos/imagenes/BOE/1939/254/A05061.tif>. Consultado el 15/02/2010.

DECRETO 2761/77, de 28 de octubre. Disposición final 4ª. En: <http://www.boe.es/boe/dias/1977/11/08/pdfs/A24430-24433.pdf>. Consultado el 15/02/2010.

Definition and purpose of a digital library. En: www.ifla.org/documents/libraries/net/alrdlib.txt. Consultado el 15/12/2009.

DEGLER, D. ; BATTLE, L. Can topic maps describe context for enterprise-wide applications? En *Extreme Markup Conference*, 2003. En: www.mulberrytech.com/Extreme/Proceedings/html/2003/Degler01/EML2003Degler01.html. Consultado el 15/12/2009.

DEPRÉS, S. ; SZULMAN, S. Construction d'une ontologie formelle à partir d'un texte de droit communautaire. *Terminology*, vol.11, 2, 2005.

DICHEVA, D. ; DICHEV, C. *Topic maps for e-learning*. En: <http://compsci.wssu.edu/iis/nsdl/index.html>. Consultado el 15/12/2009.

DIJCK, Peter van. *Introduction to XFML*. En: <http://www.xml.com/pub/a/2003/01/22/xfml.html?page=2> Consultado el 15/07/2009.

DocBook. En: www.docbook.org. Consultado el 15/12/2009.

DURUSAU, P. *Understanding Subjects and Subject Proxies*. En: http://www.durusau.net/publications/Understanding_Subjects_and_Subject_Proxies.pdf. Consultado 15/12/2009.

DURUSAU, P. *Babel and Topic Maps*. En: http://www.durusau.net/publications/Babel_and_TopicMaps.pdf. Consultado el 20/02/2010.

DURUSAU, P. ; NAITO, M. *Overview and basic concepts*. En:
C:/Users/David%20Martul/Desktop/0877.htm. Consultado el 15/12/2009.

DURUSAU, P. ; NEWCOMB, S. ; BARTA, R. *Topic maps Reference Model, 13250-5*. En:
www.isotopic-maps.org/TMRM/TMRM-7.0/tmrm7.pdf. Consultado el 15/12/2009.

E

Easy Topic maps. En: <http://easytopic-maps.com/index.php>. Consultado el 15/12/2009.

ECLA. *Euro-referencial en información y documentación. Volumen 1: competencias y aptitudes de los profesionales europeos de información y documentación*. Madrid: Sedic, 2004. En:
http://www.sedic.es/p_euro-referencial.htm. Consultado el 15/12/2009.

EITO BRUN, R. Topic Maps y la indización de recursos electrónicos en la web. *El profesional de la información*, vol.12, nº 2, 2003.

EKAW. *Knowledge Engineering and Knowledge Management by the Masses*. En:
<http://ekaw2010.inesc-id.pt/index.html>. Consultado el 20/02/2010.

ELLIS, D. A behavioural approach to information retrieval design. *Journal of Documentation*, vol.45, nº 3, 1989.

ENSER, P. ; SANDOM, C. ; HARE, J. ; LEWIS, P. Facing the reality of semantic image retrieval. *Journal of Documentation*, vol. 63, nº4, 2007.

ESCOLAR SOBRINO, H. *Historia de las bibliotecas*. Madrid : Fundación G. Sánchez Ruipérez, 1990.

F

FELLBAUM, Christiane (ed.). *Wordnet: An Electronical LexicalDatabase*. Cambridge : MIT, 1998.

FENSEL, D. *Spinning the Semantic Web: bringing the World Wide Web to its full potential*. Cambridge : MIT press, 2003.

FERRER SAPENA, A. ; PESET MANCEBO, F. ; MORENO NÚÑEZ, M. ; LLORET ROMERO, N. *Guía metodológica para la implantación de una biblioteca digital universitaria*. Gijón : Trea, 2005.

FIDEL, R. Qualitative methods in information retrieval research. *Library and Information Science Research*, vol.15, nº3. En:
<http://faculty.washington.edu/fidelf/RayaPubs/QualitativeMethodsInInformationRetrievalResearch.pdf>. Consultado el 15/02/2010.

FIDLER, R. *Mediamorfosis: comprender los nuevos medios*. Buenos Aires : Granica, 1998.

FISHER, M. Developing an information model for information- and knowledge-based organizations. En GILCHRIST, A. ; MAHON, B. *Information Architecture. Designing information environments for purpose*. London : Facet, 2005.

FONFA, R. From Faculty to Librarian Materials Selection: an element in the professionalization of librarianship. En MECH, T. F. ; MCCABE, G. B. *Leadership and Academic Libraries*. Cambridge : Greenwood Press, 1998.

FOUCAULT, M. *La arqueología del saber*. Madrid : La Piqueta, 2002.

FOULONNEAU, M. *Metadata for Digital Resources : Implementation, Systems Design and Interoperability*. Oxford : Chandos, 2008

FOURIE, I. Should we take disintermediation seriously? *The electronic library*, vol. 17, nº1, 1999.

FREESE, E. *So why aren't Topic maps ruling the world?* En:
www.idealliance.org/papers/extreme/Proceedings/html/2002/Freese01/EML2002Freese01.html. Consultado el 23/12/2009.

FREESE, E. Topic Maps vs RDF. En *Extreme Markup Language 2000*. En:
http://www.gca.org/attend/2000_conferences/Extreme_2000. Consultado el 15/12/2009.

FREESE, E. Taking Topic Maps to the Nth dimensión. En *Extreme Markup Languages* 2003. En: www.mulberrytech.com. Consultado el 15/07/2009.

G

GANGEMI, A. ; SAGRI, M. T. ; TISCORNIA, D. Metadata for content description in legal information. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies & Web based legal information management*, 2003. En: www.lri.jur.uva.nl/~winkels/LegOnt2003/Gangemi.pdf. Consultado el 12/12/2009.

GARCÍA GÓMEZ, J.C. Portales de Internet: concepto, tipología básica y desarrollo. *El Profesional de la Información*. julio-agosto, vol.10, nº 7-8. 2001.

GARCÍA JIMÉNEZ, A. *Organización y gestión del conocimiento en la comunicación*. Gijón : Trea, 2002.

GARCÍA MARTUL, D. Visualización en topic maps: tendencias y propuestas. En *JOTRI* 2003. Leganés : Departamento de Biblioteconomía y Documentación, 2003.

GARRIDO PICAZO, P. *El procesamiento automático de Documentación textual con información histórica: una aplicación XTM y DITA*. Director: Jesús Tramullas y Antonio Hernández Pérez. Universidad Carlos III de Madrid, Madrid, 2008.

GARSHOL, L. Towards a methodology for developing Topic maps Ontologies. En MAITCHER, L. ; SIGEL, A. ; GARSHOL, L. *Leveraging the Semantics of Topic maps*. Berlin : Springer, 2007.

GARSHOL, L. *Modelling with Topic maps*. En: www.topicmaps.com/tmc/presentation.jsp?conf=TM2007&id=Lars_Marius_Garshol~TM2007~Modelling_with_Topic_Maps. Consultado el 15/02/2010.

GARSHOL, L. *A citizen's portal for the city of Bergen*. En: http://www.informatik.uni-leipzig.de/~tmra/2007/slides/garshol_TMRA2007_bergen.pdf. Consultado el 5/02/2010.

GARSHOL, L. *Dtdoc*. En: <http://www.garshol.priv.no/download/software/dtdoc/>. Consultado el 15/12/2009.

GARSHOL, L. Q: A model for Topic maps-Unifying RDF and Topic maps. En *Extreme Markup Conference 2005*. En: www.mulberrytech.com/Extreme/Proceedings/html/2005/Garshol01/EMI2005Garshol01.html. Consultado el 15/12/2009.

GARSHOL, L. Towards a methodology for developing topic maps ontologies. En MAICHER, L. ; SIGEL, A. ; GARSHOL, L. *Leveraging the semantics of topic maps*. Berlin : Springer, 2007.

GARSHOL, L. *Topic maps, RDF, DAML, OIL*. En: <http://www.ontopia.net/topicmaps/materials/tmrdfoidaml.html>. Consultado el 15/12/2009.

GARSHOL, L. *Living with Topic maps and RDF*. En: <http://www.ontopia.net/topicmaps/materials/tmrdf.html>. Consultado el 15/12/2009.

GARSHOL, L. *What are Topic maps*. En: <http://www.xml.com/pub/a/2002/09/11/topicmaps.html> Consultado el 15/07/2009.

GARSHOL, L. Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. *Journal of Information Science*, vol. 30, n°4, 2004. En: www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html. Consultado el 15/12/2009.

GARSHOL, L. ; BARTA, R. *Topic Maps Query Language*. En: <http://kill.devc.at/system/files/tmq.pdf>. Consultado el 15/07/2009.

GARSHOL, L. *Modelling with Topic Maps*. En: http://www.topicmaps.com/tmc/presentation.jsp?conf=TM2007&id=Lars_Marius_Garshol~TM2007~Modelling_with_Topic_Maps Consultado el 15/07/2009.

GARSHOL, L. *The Linear Topic map Notation*. En: <http://www.ontopia.net/download/ltm.html>. Consultado el 15/12/2009.

GARSHOL, L. ; LEE, J. *GTM*. En: www.isotopic-maps.org/gtm/. Consultado el 15/12/2009.

GARSHOL, L. ; MOORE, G. *The Standard Application Model for Topic maps*. En: <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0299.htm>. Consultado el 15/12/2009.

GARSHOL, L. ; NAITO, M. *ISO 13250-4: Topic maps- Canonicalization (CXTM)*. En: <http://www.isotopic-maps.org/cxtm/>. Consultado el 15/12/2009.

GELLMAN, R. Disintermediation and the Internet. *Government Information Quarterly*, vol. 113, nº1.

GEROIMENKO, V ; CHEN, Ch. *Visualizing the Semantic Web*. London: Springer, 2005.

GIL LEIVA, I. *Manual de indización : teoría y práctica*. Gijón : Trea, 2008.

GIL URDICIÁIN, B. *Manual de lenguajes documentales*. Gijón : Trea, 2004.

GIL URDICIÁIN, B. Evaluación del rendimiento de tesauros españoles en sistemas de recuperación de información. *Revista española de documentación científica*, v.21, nº3, 1998.

GILCHRIST, A. Classification and thesauri. En VICKERY, B.C. (ed.). *Fifty years of information progress: a Journal of Documentation Review*. London: Aslib, 1994.

GILCHRIST, A. Thesauri, taxonomies and ontologies- an etymological note. *Journal of Documentation*, vol.59, nº1, 2003.

GILCRIST, A. ; KIBBY, P. ; MAHON, B. *Taxonomies for business: access and connectivity in a wired world*. London : TFPL, 2000.

GLADNEY, H. M. Digital library: Gross structure and Requirements. En *Proceedings of the First Annual Conference on the theory and practice of Digital Libraries*. Blacksburg : Virginia Polytechnic Institute, 1994.

GLAZIER, J.D. *Qualitative research in information management*. Englewood, Colorado : Libraries Unlimited , 1992.

Glosario del Portal Europa. En: http://europa.eu/scadplus/glossary/governance_es.htm
Consultado el 15/07/2009.

GNEWSYS: A kernel for semantic computing. En: <http://www.gnu.org/software/gnewsys/>.
Consultado el 15/12/2009.

GODEHARDT, E. ; BHATTI, N. Using topic maps for visually exploring various data sources in a web-based environment. En MAITCHER, L. ; GARSHOL, L. *TMR4 2007*. Berlin : Springer-Verlag, 2008. En: www.editlib.org/index.cfm/files/paper_11876.
Consultado el 15/02/2010.

GOLDFARB, Ch. *The SGML handbook*. Oxford : Clarendon Press, 1995.

GOLDFARB, Ch. Topic maps: Knowledge navigation aids. En GOLDFARB, Ch. *Charles Goldfarb's XML handbook*. New Jersey : Prentice Hall, 2002.

GÓMEZ-PÉREZ, A. ; FERNÁNDEZ-LÓPEZ, M. ; CORCHO, O. *Ontological engineering*. London : Springer, 2004.

GÓMEZ DEL PULGAR RODRÍGUEZ DE SEGOVIA, G. *Evaluación de las revistas editadas por el Centro de Estudios Políticos y Constitucionales* [memoria de licenciatura]. Getafe: Universidad Carlos III, 2000.

GÓMEZ DEL PULGAR RODRÍGUEZ DE SEGOVIA, G. *Calidad de las Revistas Españolas de Derecho Público*. [Tesis Doctoral]. Director: JOSÉ RAMÓN PÉREZ ÁLVAREZ OSSORIO. Tutor: ELÍAS SANZ CASADO. Departamento de Biblioteconomía y Documentación. Universidad Carlos III. En: <http://e-archivo.uc3m.es/bitstream/10016/495/10/Gomez%20del%20Pulgar%20Rodriguez%20de%20Segovia,%20Gloria%281%29.pdf>. Consultado el 15/02/2010.

GORMAN, G.E. *Metadata Applications and Managements*. London: Facet Publishing, 2004.

GORMAN, G. E. *The digital factor in library and information services : International Yearbook of Library and Information Management 2002-2003*. London : Facet, 2002.

GORMAN, G.E. *Qualitative Research for the information professional*. Londres : Facet publishing, 2005.

GOVERNMENT.NO. En: <http://www.regjeringen.no/en.html?id=4>. Consultado el 15/07/2009.

GRAAUW, Marc de. Business Maps: Topic maps go B2B. En *XML Europe 2001*. En: <http://www.xml.com/lpt/a/1013> . Consultado el 15/12/2009.

GRAAUW, Marc. de. *Using topic maps to extend relational databases*. En: www.xml.com/pub/a/2003/03/05/tmrdb.html. Consultado el 15/12/2009.

GRØNMO, G. *Creating semantically valid topic maps*. En: www.gca.org/papers/xml europe2000/papers/s29-02.html. Consultado el 15/12/2009.

GRUBER, T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, nº 5, 1993. En: http://ksl-web.stanford.edu//KSL_Abstracts/KSL-92-71.html. Consultado el 12/12/2009.

GUARINO, N. Formal ontology in information systems. En Guarino, N. (edit.) *I International Conference on Formal Ontology in Information Systems (FOIS'98)*. Trento : IOS Press, 1998.

GUARINO, N. ; GIARETTA, P. Ontologies and knowledge bases: towards a terminological clarification. Toward very large knowledge bases. En *Knowledge building & Knowledge Sharing*. IOS Press, 1995,

GUBAN, E. ; LINCOLN, Y. Competing paradigms in qualitative research. En DENZIN, N. ; LINCOLN, Y. (eds.). *Handbook of qualitative research*. Thousand Oaks : Sage, 1994

GUESCINI, R. ; KARABEG, D. ; NORDENG, T. A Case for Polyscopic Structuring of Information. En Maicher, L. ; Park, J. *TMR4 2005*. Leipzig : Springer, 2006.

Guidelines for the construction, format and management of monolingual controlled vocabularies. En: http://www.niso.org/kst/reports/standards/kfile_download?id%3Austring%3Aiso-8859-1=Z39-19-2005.pdf&pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVa86hjP37r_hONsJghRDv2N-zj4TZCh8Dp01rZbmK3O-8vcVjh4hezP. Consultado el 20/12/09.

GULBRANDSEN, A. Conceptual modeling of Topic maps with ORM versus UML. En *TMR4 2005*, Lecture Notes in Computer Science, vol. 3873. Leipzig : Springer, 2006. En: http://dx.doi.org/10.1007/11676904_8. Consultado el 15/02/2010.

GYUN OH, S. ; PARK, O. Design and user's evaluation of a topic maps-based Korean folk music retrieval system. En MAICHER, L. ; SIGEL, A. ; GARSHOL, L. *Leveraging the semantics of Topic Maps*. Berlin : Springer, 2007.

H

HALLIDAY, M.A.K. *El lenguaje como semiótica social: la interpretación social del lenguaje y del significado*. México : Fondo de Cultura Económica, 2001.

HAMMOND, R. The science of classification. *Professional Webmaster*, july/august 2000.

HASSAN MONTERO, Y. Visualización y Recuperación de información. En *II Encontro de Ciências e Tecnologías da Documentacao e Informacao*, Vila do Conde, 2006. En: http://www.nosolousabilidad.com/hassan/visualizacion_y_recuperacion_de_informacion.pdf. Consultado el 15/02/2010.

HATZIGAIIDAS, A. ; PAPASTERGIOU, A. ; TRYFON, G. ; MARITSA, D. Topic map existing tools: a brief review. En *Proceedings of the international conference on theory and applications of mathematics and informatics- ICTAMI 2004*. En: <http://www.emis.de/journals/AUA/acta8/Hatzigaidas-Papastergiou-Tryfon-Maritsa.pdf>. Consultado el 15/12/2009.

HAWKINS, D.T. ; LARSON, B. Q. Information Science Abstracts: Tracking the Literature of Information Science. Part 2: A new taxonomy for information science. *Journal of the American Society for Information Science and Technology*, vol. 54, n°8, 2003.

HAYNES, D. *Metadata for information management and retrieval*. London : Facet Publishing, 2004.

HEARST, M. TileBars: Visualization of term distribution information in full text information access. En *Proceedings of CHI'95*, 1995. En: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.53.538>. Consultado el 15/02/2010.

HENDLER, J. ; BERNERS-LEE, T. ; MILLER, E. Integrating applications on the Semantic Web. *Journal of the Institute of Electrical Engineers of Japan*, vol.122, 10, 2002. En: www.w3.org/2002/07/swint. Consultado el 12/12/2009.

HERNÁNDEZ GIL, A. *El cambio político español y la Constitución*. Barcelona: Planeta, 1982.

HEUER, L. ; HOPMANS, G. ; BOGACHEV, D. *ISO 13250-6: Topic maps-Compact Syntax (CTM)*. En: <http://www.isotopic maps.org/ctm/>. Consultado el 15/12/2009.

HJORLAND, B. *Information seeking and subject representation. An activity-theoretical approach to information science*. London : Greenwood Press, 1997.

HJORLAND, B. Fundamentals of knowledge organization. En *Actas del IV Coloquio Internacional de Ciencias de la Documentación y VI Congreso del Capítulo Español de ISKO*. Salamanca : Universidad de Salamanca, 2003. pp.83-116.

HÜLLEN, W. *A History of Roget's Thesaurus. Origins, Development and Design*. Oxford : Oxford University Press, 2004.

HUNTER, J.L.: A survey of Metadata Research for Organizing the Web. *Library Trends*. vol.52, n°2, 2003.

HUNTER, Ph.; GUY, M. Metadata for harvesting: the Open Archives Initiative, and how to find things on the Web. *The Electronic Library*. vol.22, n°2, 2004.

HURST, M. *About information architecture*, 2000. En: www.goodexperience.com/columns/040300infoarch.html. Consultado el 10/12/2009.

HUTHWAITE, A. AACR2 and Other Metadata Standards: The Way Forward. *Cataloging & Classification Quarterly*, vol.36, nº3/4, 2003.

HYPERGRAPH. En: <http://hypergraph.sourceforge.net/examples.html>. Consultado el 12/01/2010.

HyTime Users Group Home Page. En: <http://www.hytime.org/index.html>. Consultado el 15/12/2009.

I

IEEE Standard Glossary of Software Engineering Terminology. New York : IEEE Computer Society, 1990. En: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=159342>. Consultado el 15/02/2010.

IHMC. En: <http://www.ihmc.us/research/>. Consultado el 15/12/2009.

INGWERSEN, P. Polyrepresentation of information needs and semantic entities: Elements of a cognitive theory for information retrieval interaction. En *ACM/SIGIR-94*. London : Springer, 1994.

INGWERSEN, P. *The turn : integration of information seeking and retrieval in context*. Dordrecht : Springer, 2005.

INSALACO, T ; MASON, D. *Navigating the production maze: The Topic mapped Enterprise*. En: <http://shelter.nu/art-007.html>. Consultado el 15/12/2009.

INSTITUTE OF LAW AND TECHNOLOGY. En: <http://idt.uab.cat/> Consultado el 15/07/2009.

INTERNATIONAL HARMONIZED STAGES CODES. En: http://www.iso.org/iso/standards_development/processes_and_procedures/stages_description/stages_table.htm#s40. Consultado el 15/12/2009.

IRS Tax Map. En: <http://taxmap.ntis.gov/taxmap/>. Consultado el 15/07/2009.

ISO. Documentación. Directrices para el establecimiento y desarrollo de tesauros monolingües. Norma internacional ISO 2788-1986. *Revista Española de Documentación Científica*, 12, 4, 1989.

ISO/IEC 13250 Topic Maps [second edition]. En:

http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0322_files/iso13250-2nd-ed-v2.pdf. Consultado el 15/07/2009.

ISO/IEC FCD 18048. En:

http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?ics1=35&ics2=240&ics3=30&csnumber=52119. Consultado el 15/12/2009.

ISO/IEC FCD 19756. En:

http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?ics1=35&ics2=60&ics3=&csnumber=39987 Consultado el 15/12/2009.

iSOCO. En: http://www.isoco.com/innovacion_proyectos_nacionales.htm Consultado el 15/07/2009.

ITM T3. En:

http://mondeca.com/index.php/fr/news/archives/mondeca_announce_itm_t3

Consultado el 15/07/2009.

IUSTEL. En: <http://www.iustel.com/> Consultado el 15/12/2009.

J

JAGERMAN, E. *Creating, maintaining and applying quality taxonomies*. Zoetermeer, 2006.

JeromeDL. E-library with semantics. En: <http://www.jeromedl.org/>. Consultado el 15/12/2009.

JONES, J. *Intermediation and the digital library*. En: www.library.ubc.ca/jones/idl.html. Consultado el 15/12/2009.

JONES, D.M. ; BENCH-CAPON, T.J.M. ; VISSER, P.R.S. Methodologies for Ontology development. En *Proceedings IT&KNOWS Conference, XV IFIP World Computer Congress*. Budapest, 1998.

JUL, E. MARC and Mark-Up. *Cataloging & Classification Quarterly*, 2003, vol.36, n°3/4.

K

KARABEG, D. *Information design-a new paradigm in creation and use of information: IPSI*. En: <http://folk.uio.no/poly/ID-paradigm.pdf>. Consultado el 15/12/2009.

KARABEG, D. ; GUESCINI, R. ; NORDENG, T.W. Flexible and exploratory learning by polyscopic Topic-maps. En *ICALT 2005*. En: www.win.tue.nl/SW-EL/2005/swel05-icalt05/final/W3-2.pdf. Consultado el 12/12/2009.

KartOO. En: www.kartoo.com. Consultado el 15/12/2009.

KING, D. ; MONTGOMERY, C. H. After migration to an electronic journal collection. Impact on Faculty and Doctoral students. *D-Lib Magazine*, vol.8, n°12, 2002. En: <http://www.dlib.org/dlib/december02/king/12king.html>. Consultado el 15/02/2010.

KING, B. ; REINOLD, K. *Finding the concept, not just the word: a librarian's guide to ontologies and semantics*. Oxford : Chandos Publishing, 2008.

KLEIN, J. T. *Interdisciplinarity: History, theory and practice*. Detroit : Wayne State University Press, 1990.

KNAPP, A. *La experiencia del usuario*. Madrid : Anaya Multimedia, 2002.

KRUK, S. *Semantic Digital Libraries. Improving usability of information discovery with semantic and social services*. Lulu.com, 2010.

KRUK, S. R. [et al.] *Tutorial on semantic digital libraries at ESWC'2007*. En: <http://library.deri.ie/resource/goIjrewV>. Consultado el 15/12/2009.

KRUK, S. ; MCDANIEL, B. *Semantic Digital Libraries*. Berlin : Springer, 2009.

KUHN, T. *La estructura de las revoluciones científicas*. México: Fondo de Cultura Económica, 2001.

L

LAGOZE, C. ; PAYETTE, S. ; SHIN, E. ; WILPER, C. Fedora: an architecture for complex objects and their relationships. *International Journal on Digital Libraries*. Vol. 6, nº 2, 2006.

LAMARCA LAPUENTE, M.J. *Hipertexto: El nuevo concepto de documento en la cultura de la imagen*. En: http://www.hipertexto.info/documentos/maps_tematic.htm. Consultado el 15/12/2009.

LAME, G. Using text analysis techniques to identify legal ontologies components. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies & Web based legal information management*, 2003. En: <http://www.lri.jur.uva.nl/~winkels/LegOnt2003/Lame.pdf> Consultado el 12/12/2009.

LANCASTER, F.W. *Indexing and abstracting in theory and practice*. London : The library Association, 1991.

LANCASTER, F.W. *El control del vocabulario en la recuperación de la información*. Valencia : Universitat de Valencia, 2002.

LANDOW, P. *Teoría del hipertexto*. Barcelona : Paidós, 1997.

LANGRIDGE, D. W. *Subject analysis: Principles and procedures*. London : Bowker-Saur, 1989.

LARSBLOG. En: <http://www.garshol.priv.no/blog>. Consultado el 15/12/2009.

LAUREL, B. *The Art of Human Computer Interface Design*. Reading : Addison-Wesley, 1999.

LAVIK, S. ; NORDENG, M. BrainBank Learning. Building topic maps as strategy for learning. En *XML Europe 2004*. En: www.idealliance.org/papers/dx_xmle04/papers/04-02-03/04-02-03.html. Consultado el 15/12/2009.

LAZAR, J. *Web usability. A user-centered design approach*. Boston : Pearson, 2005.

LE GRAND, B. Topic map Visualization. En Park, J. *XML Topic maps*. Boston : Addison-Wesley, 2002.

LE GRAND, B. ; SOTO, M. Topic maps metrics. En *Proceedings of Knowledge Technologies 2001*. En: <http://www2.gca.org/knowledgetechnologies/2001/proceedings/LeGrand%20Slides.pdf> . Consultado el 15/12/2009.

LEARY, R. M. ; VANDENBERGHE, W. ; ZELEZNIKOW, J. Towards a financial fraud ontology. A legal modeling approach. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies & Web based legal information management*, 2003. En: <http://www.lri.jur.uva.nl/~winkels/LegOnt2003/Leary.pdf> Consultado el 12/12/2009.

LegOnt 2001. En *Proceedings of the Second International Workshop on legal ontologies*. Amsterdam, 2001. En: www.lri.jur.uva.nl/jurix2001/legont2001.htm. Consultado el 12/12/2009.

LegOnt 2003. En *Proceedings of the ICAIL 2003 Workshop on Legal Ontologies & Web based legal information management*. Edinburgh, 2003. En: www.lri.jur.uva.nl/~winkels/legontICAIL2003.html. Consultado el 12/12/2009.

LegalXML. En: <http://www.legalxml.org>. Consultado el 16/12/2009.

LEHMAN, J.; BREUKER, J. ; BROWER, B. Causation in AI&Law. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies & Web based legal information management*, 2003. En: <http://www.lri.jur.uva.nl/~winkels/LegOnt2003/Lehmann.pdf>. Consultado el 12/12/2009.

LEXML. En: <http://www.uv.es/lexml>. Consultado el 15/12/2009.

LÍ, V. ; MASON, D. Topic maps for managing classification guidance. En *Proceedings of Extreme Markup Languages*, 2002. En: www.mulberrytech.com/Extreme/Proceedings/html/2002/Mason01/EML2002Mason01-toc.html. Consultado el 15/05/2009.

LIBRELOTTO, G. ; RAMALHO, E. ; HENRIQUES, P. *XTche*. En: <https://repositorium.sdum.uminho.pt/retrieve/1358/XML2004-XTche.pdf>. Consultado el 15/12/2009.

LIU, D. ; HSU, CH. *Project-based knowledge maps: combining project mining and XML-enabled topic maps*. En: www.emeraldinsight.com/Insight/ViewContentServlet?Filename=Published/EmeraldFullTextArticle/Articles/1720140306.html. Consultado el 15/02/2010.

LLORENS, J. *Definición de una metodología y una estructura de repositorio orientadas a la reutilización : el tesoro de software*. Director: Antonio AMESCUA SECO. Universidad Carlos III de Madrid, Departamento de Informática, 1996.

LOPEZ ALONSO, M.A. Evolución histórica y tendencias observables en los tesauros. *Boletín Millares Carlo*, nº17, 1998.

LÓPEZ HUERTAS, M.J. Thesaurus structure design: a conceptual approach for improved interaction. *Journal of Documentation*, vol. 53, nº 2, 1997.

LÓPEZ YEPES, J. El nuevo profesional de la información, del conocimiento y de la comunicación: El bibliotecario universitario. *Anales de Documentación*, nº10, 2007. En: <http://revistas.um.es/documentacion/article/viewFile/1191/124>. Consultado el 15/12/2009.

LOZANO TELLO, A. Ontologías en la web semántica. *Cuadernos de investigación en ingeniería informática*, nº5, 2001. En: www.informandote.com/jornadasIngWEB/articulos/jiw02.pdf. Consultado el 14/12/2009.

LU, Y. ; HORNUNG, J. *3D visualization of knowledge domains in learning environments*. En: www.editlib.org/index.cfm/files/paper_11876.pdf. Consultado el 15/02/2010.

LYMAN, P. *How much information*. En:

www.sims.berkeley.edu/research/projects/howmuch-info-2003/ Consultado el 12/12/2009.

M

MACIÁ, M. *Manual de documentación jurídica*. Madrid : Síntesis, 1998.

MAITCHER, L. ; WITSCHER, H. F. *Merging of distributed Topic maps based on the Subject Identity Measure (SIM)*. En: www.lutzmaicher.de/PUBLIKATIONEN. Consultado el 15/12/2009.

MARCHIONINI, G. *Information seeking in electronic environments*. Cambridge : Cambridge University Press, 1997.

MARCHIONINI, G. Búsqueda exploratoria: de encontrar a comprender. *Novática: Revista de la Asociación de Técnicos de Informática*, nº185, 2006.

MARCOS, M.C. Elementos visuales en sistemas de búsqueda y recuperación de información. *Hipertext.net*, nº3, 2005. En: <http://www.hipertext.net/web/pag257.htm#13338>. Consultado el 15/12/2009.

MARCOS, M. C. KartOO ahora más español. *El Profesional de la Información*, vol.12, nº2, 2003. En: <http://www.elprofesionaldelainformacion.com/contenidos/2003/marzo/10.pdf>. Consultado el 15/12/2009.

MASON, J. *Topic maps –XML Syntax*. En: <http://www.isotopic-maps.org/sam/sam-xtm/>. Consultado el 15/12/2009.

MAZZOCCHI, F. [et al.]. Relational Semantics in Thesauri: Some remarks at theoretical and practical levels. *Knowledge Organization*, vol.34, nº4, 2007.

McCALLUM, S.H. An introduction to the Metadata Object Description Schema (MODS). *Library Hi Tech*, vol.22, nº1, 2004.

MCCARTY, L.T. A language for legal discourse, I. Basic Features. En *Proceedings of the second international conference on artificial intelligence and law*. Vancouver, 1989.

MCCOMBS, G.M. ; MAYLONE, T.M. Qualitative research. [*Special Issue*]. *Library Trends*, vol.46, nº4, 1998.

MANIEZ, J. *Actualité des langages documentaires : les fondements théoriques de la recherche d'information*. Paris : Association des professionnels de l'information et de la documentation (ADBS), 2002.

Manifiesto UNESCO/IFLA sobre la Biblioteca Escolar. En:

http://www.unesco.org/webworld/libraries/manifestos/school_manifesto_es.html

Consultado el 04/12/2009.

MARCOS, M.C. *Interacción en interfaces de recuperación de información: conceptos, metáforas y visualización*. Gijón : TREA, 2004.

MARTÍN GALÁN, B. ; RODRÍGUEZ MATEOS, D. Estructuración de la información mediante XML: un nuevo reto para la gestión documental. En *VII Jornadas Españolas de Documentación*. Bilbao: Universidad del País Vasco, 2000.

MARTÍNEZ EQUIHUA, S. *Biblioteca digital. Conceptos, recursos y estándares*. Buenos Aires : Alfagrama, 2007.

MEGHINI, C. ; SPYRATOS, N. Information Access in digital libraries: Steps towards a conceptual schema. En IOANNIDIS, Y. ; SCHEK, H.J.; WEIKUM, G. *Future Digital Library Management Systems: System Architecture and Information Access*. Munich : DELOS, 2005.
En: http://www.delos.info/files/pdf/Proceedings/Dagstuhl_2903_010405/delos-dagstuhl-handout-all.pdf. Consultado el 13/12/2009.

MEMORIA 2008. *Centro de Estudios Políticos y Constitucionales*. En:

www.cepc.es/include_mav/getfile.asp?IdFileImage=3241. Consultado el 15/02/2010.

MÉNDEZ RODRÍGUEZ, E. *Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales*. Gijón : Trea, 2002.

MODELS (*Moving to distributed environments for library services*). En:

<http://www.ukoln.ac.uk/dlis/models/requirements/overview/>. Consultado el 15/12/2009.

MOMMERS, L. *Applied legal epistemology. Building a knowledge-based ontology of the legal domain*. Leiden, 2003. En:

https://openaccess.leidenuniv.nl/dspace/bitstream/1887/4432/1/Proefschrift_Mommers.pdf. Consultado el 25/12/2009.

MONDECA. En: www.mondeca.com. Consultado el 15/07/2009.

MOOERS, CALVIN N. Zatocoding applied to mechanical organization of knowledge. *American Documentation*, nº 2, 1951.

MOORE, G. Topic map technology. State of the art. En *XML Europa 2000*. En: www.gca.org/papers/xml europe2000/papers/s22-04.html. Consultado el 15/12/2009.

MOORE, G. *RDF and Topic maps: an exercise in convergence*. En: www.topic-maps.com/topic-mapsrdf.pdf. Consultado el 14/12/2009.

MORALES CAMPOS, E. *La biblioteca del futuro*. México : UNAM, 1996.

MORATO, J; MARZAL, M.A.; LLORENS, J; MOREIRO, J. WordNet applications. En SOJKA, P.; PALA, K. ; SMRZ, P. ; FELLBAUM, CH. ; VOSSEN, P. *GW C 2004: Proceedings of the Second International WordNet Conference*. Brno : Masaryk University, 2004.

MOREIRO GONZÁLEZ, J.A. *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural*. Gijón : Trea, 2004.

MOREIRO GONZÁLEZ, J. A. La representación y recuperación de los contenidos digitales: De los tesauros conceptuales a las folksonomías. En TRAMULLAS, J. *Tendencias en documentación digital*. Gijón : Trea, 2006.

MOREIRO GONZÁLEZ, J.A.[et al.]. Desarrollo profesional y opinión sobre la formación recibida de los titulados universitarios en información y documentación de las universidades públicas de Madrid (2000-2005). *El Profesional de la Información*, vol.17, nº3.

MOREIRO GONZÁLEZ, J.A.[et al.]. Metodología del proyecto REID : Desarrollo de un sistema de recuperación para entornos de información dinámica: Tesauros de verbos, implementación del estándar ISO/ICE 13250:1999. En *Jornadas de Tratamiento y Recuperación de Información*. En: www.fundacion.uc3m.es/jotri2003/ponencias/metodologia.pdf. Consultado el 15/12/2009.

MOREIRO GONZÁLEZ, J.A. ; LLORÉNS MORILLO, J. ; MARZAL GARCÍA-QUISMONDO, M.A. Mapas Conceptuales, *Topic maps* y Tesauros. En *Jornadas de Tratamiento y Recuperación de la Información*. Valencia : Universitat de Valencia, 2002.

MOREIRO, J. A. ; SÁNCHEZ CUADRADO, S. ; MORATO LARA, J. Panorámica y tendencias en topic maps. *Anuario Hipertext.net*, nº1, Mayo 2003. En: <http://www.hipertext.net/web/pag229.htm>. Consultado el 15/12/2009.

MORIYA, A. ; LOPES M. da SILVA, F. ; MAGALHAES JR, W. *Topic Maps*. Director: Maria Luiza Machado Campos. [Tesis no publicada]. Universidade Federal do Rio de Janeiro, Instituto de Matemática, 2003.

MORVILLE, P. ; ROSENFELD, L. *Information Architecture*. Beijing : O'Reilly, 2006.

MORVILLE, P. A brief history of information architecture. En GILCHRIST, A. ; MAHON, B. *Information Architecture: designing information environments for purpose*. London: Facet, 2004.

MUGNAINI, L. Mapping Topic maps on relational databases. En *XML Europe, 2000*. En: www.gca.org/papers/xml europe2000/papers/s22-04.html. Consultado el 15/12/2009.

MUNZNER, T. H3: Laying Out Large Directed Graphs in 3D Hyperbolic Space. En *Proceedings of the 1997 IEEE Symposium on Information Visualization*, 1997. En: <http://www.graphics.stanford.edu/papers/h3/>. Consultado el 15/12/2009.

N

NAITO, M. ; PEPPER, S. *Draft 13250-1 Topic maps—Overview and Basic Concepts*. En: www.jtc1sc34.org/repository/0877.zip. Consultado el 15/12/2009.

NATIONAL LIBRARY OF AUSTRALIA. En: <http://www.nla.gov.au/find/books.html>. Consultado el 15/12/2009.

NEUMAN, D. High school student's use of databases: results of a national Delphi study. *Journal of the American Society for Information Science*, vol. 46, n°4.

NEW ZEALAND ELECTRONIC TEXT CENTRE (NZETC). En: <http://www.nzetc.org>. Consultado el 15/05/2009.

NEWCOMB, S. A perspective on the quest for global knowledge interchange. En PARK, J. *XML Topic Maps*. Boston : Addison-Wesley, 2003.

NEWCOMB, S. *A semantic integration methodology*. En: http://www.idealliance.org/papers/extreme/Proceedings/html/2003/Newcomb01/EML_2003Newcomb01.html. Consultado el 23/12/2009.

NEWCOMB, S. Standards. Standard Music Description Language Complies with Hypermedia Standard. *IEEE Computer* 24/7, 1991.

NEWCOMB, S. ; BIEZUNSKI, M. ; BRYAN, M. *The HyTime Topic maps (HyTM) syntax 1.0*. En: <http://www.jtc1sc34.org/repository/0391.htm>. Consultado el 15/12/2009.

NIEDERHAUSEN, H. *Creating a topic map ontology with Onotoa*. En: <http://onotoa.topicmapslab.de/tutorial>. Consultado el 15/12/2009.

NIELSEN, J. *Usabilidad. Diseño de sitios web*. Madrid: Pearson Educación, 2001.

NISHIKAWA, M. Organizing Information in a Corporate Intranet: A Use Case for Published and Internal-Use Subjects in Topic Maps. En *Proceedings of Extreme Markup Languages 2002*. En:

<http://conferences.idealliance.org/extreme/html/2002/Nishikawa01/EMI.2002Nishikawa01.html>. Consultado el 15/12/2009.

NOGALES FLORES, T. ; MARTÍN GALÁN, B. ; ARELLANO PARDO, M.C. Informática, Derecho y Documentación. Experiencias y posibilidades de aplicación de los lenguajes de marcado de texto (SGML, HTML y XML) a los documentos jurídicos. En *Encuentro sobre Informática y Derecho*. Madrid : Instituto de informática jurídica, 2002. En: <http://www.bib.uc3m.es/~bmartin/publicaciones/icade2002.pdf>. Consultado el 15/02/2010.

NOGALES FLORES, T. ; MARTÍN GALÁN, B. *Difusión y recuperación en la Web de documentos jurisprudenciales haciendo uso de la tecnología XML*. En: www.bibliociencias.cu/gsd/collect/eventos/index/assoc/HASH0159.dir/doc.pdf. Consultado el 10/02/2010.

NOTICIAS JURÍDICAS. En: <http://noticias.juridicas.com/>. Consultado el 15/12/2009.

NORDENG, T. ; GUESCINI, R. ; KARABEG, D. Topic maps for polyscopic structuring of information. *International Journal of Continuing Engineering Education and Lifelong Learning*, vol. 16, nº1/2, 2006.

Norm@Civil. En: <http://civil.udg.es/normacivil/> Consultado el 15/12/2009.

NOVAK, J. D. ; CAÑAS, A. *The theory underlying concept maps and how to construct and use them*. En: <http://cmap.ihmc.us/Publications/ResearchPapers/TheoryCmaps/TheoryUnderlyingConceptMaps.htm>. Consultado el 15/12/2009.

NOVAK, J. D. ; GOWIN, D. B. *Learning how to learn*. New York : Cambridge University Press, 1984.

NOY, N. ; McGUINNES, D. *Ontology Development 101: A Guide to Creating Your First Ontology*. En:

http://protege.stanford.edu/publications/ontology_development/ontology101.pdf.

Consultado el 15/02/2010.

NOY, N. ; McGUINNES, D. *Desarrollo de ontologías-101: Guía práctica para crear tu propia ontología*. En:

http://protege.stanford.edu/publications/ontology_development/ontology101-es.pdf.

Consultado el 20/02/2010.

O

OAKRIDGE NATIONAL LABORATORY. En: <http://www.ornl.gov/>. Consultado el 15/07/2009.

OASIS. En: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=geolang
Consultado el 15/07/2009.

OBRST, L. ; LIU, H. Knowledge Representation, Ontological Engineering, and Topic maps. En PARK, J. *XML Topic maps. Creating and Using Topic maps for the Web*. Boston: Addison-Wesley, 2002.

OGIEVETSKY, N. Building dynamic web sites with topic maps and XSLT. En *Extreme Markup Conference 2000*. En: <http://ep.open.ac.uk/PubSys/resources/html/ogie1026.html>.
Consultado 15/12/2009.

OGIEVETSKY, N. ; BADGER, T. Topic map solutions for Kodak digital camera accessories. En *XML Europe 2003*. En:
http://www.idealliance.org/papers/dx_xmle03/papers/02-03-02/02-03-02.pdf .
Consultado el 15/12/2009.

OGIEVETSKY, B. Tabular Topic maps. En *XML Europe 2003*. En: <http://web-services.gov/TTMSBIR20031016.ppt>. Consultado el 15/12/2009.

OLIVEIRA, A. Topic Maps na visualização de informação no ensino e treino. En *Proceedings da Associação Portuguesa de Sistemas de Informação*. Guimarães : Conferência da

Associação Portuguesa de Sistemas de Informação, 2001. En:
<http://en.scientificcommons.org/20329751>. Consultado el 15/02/2010.

Omnigator: The Topic Map browser. User's guide. En:
<http://www.ontopedia.net/omnigator/docs/navigator/userguide.html>. Consultado el 21/07/2009.

ONOTOA. En: <http://onotoa.topic-mapslab.de>. Consultado el 15/12/2009.

ONTOPOLY. En: <http://www.ontopia.net/solutions/ontopoly.html>. Consultado el 15/10/2009.

ORDEN de 21 de julio de 1966. Reglamento orgánico. BOE nº191 de 11/08/66. En:
<http://www.boe.es/datos/pdfs/BOE/1966/191/A10466-10469.pdf>. Consultado el 15/02/2010.

OROZCO GÓMEZ, G. *La investigación en comunicación desde la perspectiva cualitativa*. La Plata: Facultad de Periodismo y Comunicación social, Universidad Nacional de la Plata; Guadalajara : Instituto Mexicano para el desarrollo comunitario, 1997.

ORTEGA Y GASSET, J. Misión del bibliotecario. *Revista de Occidente*, año XIII, nº CXLIII, 1935.

ORTIZ-REPISO, V. [et al.]. How researchers are using the OPAC of the Spanish Council for Scientific Research Library Network. *Electronic Library*, vol. 24, nº 2, 2006.

ÖZEL, S.A. ; ALTINGÖVDE, S. ; ULUSOY, O. ; ÖZSOYOĞLU, G. ; ÖZSOYOĞLU, Z.M. Metadata-Based Modeling of Information Resources on the Web. *Journal of the Information Society for Information Science and Technology*, vol.55, nº2, 2004.

P

PARK, J. *XML Topic maps. Creating and using Topic maps for the Web*. Addison-Wesley, 2003.

PASSIN, T. *Explorer's Guide to the Semantic Web*. Greenwich : Manning, 2004.

PASSIN, T. Browser Bookmark Management with Topic maps. En *Extreme Markup Conference 2003*. En: www.mulberrytech.com/Extreme/Proceedings/html/2003/Passin01/EML2003Passin01.html. Consultado el 15/12/2009.

PASSIN, T. Navigating information with topic maps. En PASSIN, T. *Explorer's guide to the semantic web*. Greenwich : Manning, 2004.

PEARCE, J. ; CATHRO, W. ; BOSTON, T. The challenge of integrated access: the hybrid library system of the future. En *10º VALA Biennial Conference and Exhibition*. Melbourne, 2000. En: www.nla.gov.au/nla/staffpaper/jpearce1.html. Consultado el 15/12/2009.

PEPPER, S. The Italian Opera Topic map. En: <http://www.ontopia.net/operamap/index.jsp>. Consultado el 15/12/2009.

PEPPER, S. *The TAO of Topic Maps. Finding the way in the age of infoglut*. En: <http://www.ontopia.net/topicmaps/materials/tao.html>. Consultado el 12/02/2010.

PEPPER, S. *The case for published subjects*. En: <http://www.ontopia.net/topicmaps/materials/The Case for Published Subjects.pdf>. Consultado el 12/02/2010.

PEPPER, S. *Published Subjects: Introduction and Basic Requirements*. En: <http://www.oasisopen.org/committees/download.php/2897/pubsub-pt1-1.01-cs.pdf>. Consultado el 18/02/2010.

PEPPER, S. *Mythesaurus*. En: http://localhost:8080/omnigator/models/topicmap_complete.jsp?tm=MyThesaurus.xtm. Consultado el 15/01/2010.

PEPPER, S. Topic Maps. En BATES, M. ; MAACK, M. *Encyclopedia of Library and Information Sciences*. Boca Raton, FL : CRC Press, 2009.

PEPPER, S. *Towards Seamless Knowledge. Integrating Public Sector Portals in Norway*. En: www.ontopia.net/topicmaps/materials/Towards%20Seamless%20Knowledge.ppt. Consultado el 15/12/2009.

PEPPER, S. *The case for published subjects*. En:

http://www.ontopia.net/topicmaps/materials/The_Case_for_Published_Subjects.pdf.

Consultado el 10/02/2010.

PEPPER, S. Expressing Dublin Core in Topic Maps. En MATTCHEER, L. ; GARSHOL, L. *Scaling Topic Maps. Third International Conference on Topic Maps Research and Applications, TMR4 2007*. Berlin : Springer, 2007.

PEPPER, S. *Creating a topic map*. En: www.topicmaps.com/tm2008/pepper3.pdf.

Consultado el 15/02/2010.

PEPPER, S. Navigating haystacks and discovering needles. Introducing the new topic map standard. En *Markup Languages: Theory & Practice*, 1999. En: <http://www.ontopia.net/topicmaps/materials/mlangart.pdf>. Consultado el 15/12/2009.

PEPPER, S. ; MOORE, G. *XML Topic maps (XTM) 1.0*. En: <http://www.topicmaps.org/xtm/1.0/index.html>. Consultado el 15/12/2009.

PEPPER, S. ; GRONMO, G. O. *Towards a General Theory of Scope*. En:

<http://www.ontopia.net/topicmaps/materials/scope.htm>. Consultado el 15/07/2009.

PEPPER, S. ; SCHWAB, S. *Curing the Web's Identity Crisis*. En:

<http://www.ontopia.net/topicmaps/materials/identitycrisis.html> Consultado el

15/07/2009.

PETRUCK, M. Frame Semantics. En VERSCHUEREN, J. ; ÖSTMAN, J. ; BLOMMAERT, J. ; BULCAEN, Ch. (eds.). *Handbook of Pragmatics*. Philadelphia : John Benjamins, 1996.

PFITZNER, D. ; HOBBS, V. ; POWERS, D. A Unified Taxonomic Framework for Information Visualization. En PATTISON, T. ; THOMAS, B. (eds.). En *Proceedings Australian Symposium on Information Visualisation*. Adelaide : Conferences in Research and Practice in Information Technology, 2001. En:

<http://crpit.com/confpapers/CRPITV24Pfitzner.pdf>. Consultado el 15/02/2010.

PICKARD, A. J. *Research Methods in Information*. London : Facet Publishing, 2007.

PINK, D. H. *A whole new mind. Moving from the information age to the conceptual age*. New York : Riverhead books, 2005.

PLAZ LANDAETA, R. ; GONZÁLEZ AURE, N. *La gestión del conocimiento organizativo. Dinámicas de agregación de valor en la organización*. En:
www.mityc.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/357/05_ReinaldoPlaz_357.pdf. Consultado el 10/02/2010.

R

RATH, H. *Topic map Fundamentals for Knowledge Representation*. En PARK, J. XML *Topic maps*. Boston : Addison-Wesley, 2002.

RATH, H. *Making Topic maps more colourful*. En:
www.gca.org/papers/xmleurope2000/papers/s29-01.html. Consultado el 15/12/2009.

RATH, H. *Semantic Resource Exploration with Topic maps*. En: www.uni-giessen.de/germanistik/ascl/gldv2001/proceedings/pdf/GLDV2001-rath.pdf. Consultado el 15/12/2009.

RATH, H. *The Topic maps Handbook*. Gütersloh : Empolis, 2003. En:
http://www.empolis.com/download/docs/whitepapers/empolistopicmapswitepaper_eng.pdf. Consultado el 15/12/2009.

RATH, H. ; MOORE, G. *Topic map Constraint Language (TMCL) Requirements and Use Cases*. En: <http://www.isotopicmaps.org/tmcl/requirements.html>. Consultado el 15/12/2009.

RATH, H. ; PEPPER, S. *Topic maps: Introduction and Allegro*. En: www.topicmaps.com/content/resources/mt99/hhr-stp.pdf. Consultado el 15/12/2009.

RATH, H. ; PEPPER, S. *Topic maps at work*. En GOLDFARB, C.F. ; PRESCOND, P. (eds.). *XML handbook*. Englewood Cliffs, NJ: Prentice Hall, 1999.

RDF/Topic maps Interoperability Task Force. En:

<http://www.w3.org/2001/sw/BestPractices/RDFTM>. Consultado el 15/12/2009.

REAL WORLD APPLICATIONS OF TOPIC MAPS. En:

<http://www.idealliance.org/papers/xml2001papers/tm> . Consultado el 15/12/2009.

REAL DECRETO 1707/80, de 29 de agosto de 1980. Artº1º. En:

<http://www.boe.es/boe/dias/1980/09/04/pdfs/A19942-19943.pdf>. Consultado el 15/02/2010.

REAL DECRETO 1269/1997, de 7 de agosto de 1997, de organización y funcionamiento del Centro de Estudios Políticos y Constitucionales (BOE nº214). En:

<http://www.boe.es/boe/dias/1997/08/07/pdfs/A24078-24080.pdf>. Consultado 15/02/2010.

REDERO SAN ROMÁN, M. *La transición a la democracia en España*. Madrid : Marcial Pons, 1996.

REY, A. *La terminologie: noms et notions*. Paris : Presses Universitaires de France, 1979.

RM4TM. En: www.isotopic maps.org/rm4tm. Consultado el 15/12/2009.

RITTERSHOFER, A. *Lerner mit Topic maps*. En: <http://www.lmtm.de> . Consultado el 15/12/2009.

ROCHA OLIVEIRA, A. *Deployment of Topic Maps for Navigation and Searching in Huge Information Spaces as Component of Learning Environment*. Braga : Centro de Computacao Gráfica, 2000. En:

http://www.ccg.pt/Publications/PDFs/Theses/2000/Diploma_Thesis.pdf. Consultado el 15/02/2009.

RODRÍGUEZ BRAVO, B. *El documento: entre la tradición y la renovación*. Gijón : Trea, 2002.

ROSA IGLESIAS, B. El “Centro de Estudios Políticos y Constitucionales” de Madrid. *Historia Constitucional*, nº3, 2002. En:

<http://www.historiaconstitucional.com/index.php/historiaconstitucional/article/view/187/165>. Consultado el 15/02/2010.

ROSENFELD, L. ; MORVILLE, P. *Information Architecture for the World Wide Web*. USA : O'Reilly, 2002.

ROVIRA FONTANALS, C. El editor de mapas conceptuales DIGIDOCMAP y la norma topic maps. *Hipertext.net*, nº3, 2005. En: <http://dialnet.unirioja.es/servlet/articulo?codigo=1176266>. Consultado el 15/12/2009.

ROVIRA FONTANALS, C. ; MARCOS, M.C. ; CODINA, Ll. Repositorios de publicaciones digitales de libre acceso en Europa: análisis y valoración de la accesibilidad, posicionamiento web y calidad del código. *El profesional de la información*, vol. 16, nº 1, 2007. En: http://www.mcmarcos.com/pdf/2007-repositorios_EPI.pdf. Consultado el 15/02/2010.

ROWLEY, J. *Organizing knowledge: an introduction to managing access to information*. Aldershot: Gower Publishing Limited, 2000.

ROWLEY, J. *Organizing Knowledge: an introduction to managing access to information*. Hampshire : Ashgate, 2008.

RUGGLES, Rudy L., (ed). *Knowledge management tools*. Boston : Butterworth-Heinemann, 1997.

RYSZARD KRUK, S. *Semantic Digital Libraries. Improving usability of information discovery with semantic and social services*. Dublin, 2010.

S

SAIAS, J. ; QUARESMA, P. Using NLP techniques to create legal ontologies in a logic programming based web information retrieval system. En *Proceedings of the ICAIL 2003 Workshop on legal ontologies & Web based legal information management*, 2003. En: <http://www.lri.jur.uva.nl/~winkels/LegOnt2003/Saias.pdf>. Consultado el 12/12/2009.

SANZ CASADO, E. *Manual de estudios de usuarios*. Madrid : Fundación Germán Sánchez Ruipérez, 1994.

SARACEVIC, T. *Digital Library: toward an evolution of concepts*. En: https://docubib.uc3m.es/BIBRECOMENDADA/GESTDOC/gestdoc_06.pdf. Consultado 15/12/2009.

SARACEVIC, T. Evaluation of evaluation in information retrieval. En *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*. New York : ACM, 1995. En: <http://delivery.acm.org/>. Consultado 15/12/2009.

SAUSSURE, F. *Curso de Lingüística General*. Bueno Aires : Losada, 2002.

SCHWEIGHOFER, E. The revolution in legal information retrieval or the Empire strikes back. *Journal of Information, Law and Technology (JILT)*. En: <http://elj.warwick.ac.uk/jilt/99-1/schweigh.html>. Consultado el 15/12/2009.

SEIDL, F. Elmer Preview: A standards-based architecture for creating a unified information space. En *XML 2001 Conference*. En: www.idealliance.org/papers. Consultado el 15/07/2009.

SEMANTEXT. En: <http://sourceforge.net/projects/semantext/>. Consultado el 15/12/2009.

SENSO, J.A. El concepto de metadato. Algo más que descripción de recursos electrónicos. *Ciência da Informação*, vol.32, nº2, 2003.

SERRANO-COBO, J. Combinación de logs internos y externos en la predicción de estacionalidad de búsquedas para el rediseño de webs. *El Profesional de la Información*, vol.18, nº1, enero-febrero 2009.

SHANNON, C. A mathematical theory of communication. *Bell system technical journal*, july 1948. En: <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>. Consultado el 09/01/2010.

SIERRA BRAVO, R. *Tesis Doctorales y trabajos de investigación científica*. Madrid : Thomson, 2005.

SIGEL, A. Topic maps in knowledge organization. En PARK, J. *XML Topic Maps*. Boston : Addison-Wesley, 2003.

SIGEL, A. Topic maps in Knowledge Organization. En PARK, J. ; HUNTING, S. *XML Topic maps*. Boston : Addison-Wesley, 2002.

SIGEL, A. Towards Knowledge Organization with Topic maps. En *XML Europe2000*, 2001. En: <http://www.gca.org/papers/xmleurope2000/papers/s22-02.html>. Consultado el 09/02/2010.

SILIÓ, T. *La isomorfía lingüística sobre la base de la iteración recursiva, la coherencia estructural y la autosemejanza del lenguaje*. Director: Tomás Albaladejo Mayordomo. Universidad Autónoma de Madrid, Madrid, 2004.

SIMPLE KNOWLEDGE ORGANIZATION SYSTEM. En: www.w3.org/2004/02/skos/. Consultado el 15/05/2009.

SLAVIC, A. ; BAIGET, C. Using Dublin Core in educational material: some practical considerations based on the EASEL experience. *VINE*, nº 125.

SLYPE, G.van. *Lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Peñaranda : Fundación Germán Sánchez Ruipérez, 1991.

SMITH, B. *Ontology*, 2003. En: http://ontology.buffalo.edu/smith/articles/ontology_pic.pdf. Consultado el 12/12/2009.

SMOLNIK, S. ; NASTANSKY, L. *K-Discovery: Identification of Distributed Knowledge Structures in a Process-oriented Groupware Environment*, 2001. En: <http://www2.gca.org/knowledgetechnologies/2001/proceedings/Smolnik%20Paper.pdf> Consultado el 15/12/2009.

SMOLNIK, S. ; NASTANSKY, L. K-Discovery: Using Topic maps to identify Distributed Knowledge Structures in Groupware based Organizational Memories. En *Proceedings of the 35th Hawaii International Conference on System Sciences*, 2002. En:

<http://www2.computer.org/portal/web/csdl/abs/proceedings/hicss/2002/1435/04/14350106babs.htm>. Consultado el 15/12/2009.

SOERGEL, D. *Organizing information: Principles of data base and retrieval systems*. London : Academic Press, 1985.

SOLÉ TURA, J. *Constituciones y periodos constituyentes en España (1808-1936)*. Madrid : Siglo XXI, 1990.

SOWA, J. *Conceptual Structures*. Reading : Addison-Wesley, 1984.

SOWA, John F. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. New York : Brooks Cole, 2000.

SOWA, J. *Building, Sharing, and Merging Ontologies*. En:

<http://www.jfsowa.com/ontology/ontoshar.htm>. Consultado el 19/02/2010.

SPYNS, P. ; DE BO, J. Ontologies: a revamped cross-disciplinary buzzword or a truly promising interdisciplinary research topic? *Lingüística Antverpiensia. New Series* 3, 2004.

STUDER, S. ; BENJAMINS, R. ; FENSEL, D. Knowledge engineering: principles and methods. *Data and Knowledge Engineering*, 25, 1998. En:

<http://citeseer.ist.psu.edu/225099.html>. Consultado el 12/12/2009.

SUTTON, B. The rationale for qualitative research: a review of principles and theoretical foundations. *Library Quarterly*, vol.63, nº4, 1993.

SVENONIUS, E. *The intelectual foundation of information organization*. Cambridge: MIT Press, 2000.

T

TAYLOR, S. J. ; BOGDAN, R. *Introducción a los métodos cualitativos de investigación*. Barcelona : Paidós, 1994.

TAYLOR, A. G. *The organization of information*. London : Libraries unlimited, 2004.

The www Virtual Library. En: <http://vlib.org/index.en>. Consultado 04/12/2009.

Techquila shots. En: <http://www.techquila.com/blog>. Consultado el 15/12/2009.

TEMMERMAN, R. ; KNOPS, U. Introduction. The translation of domain specific languages and multilingual terminology management. *Lingüística Antverpiensia New Series*, 3, 2004.

TEMMERMAN, R. ; KERREMANS, K. Termonography: Ontology building and the sociocognitive approach to terminology description. En Hajicová, E. ; Kotesovcová, A. ; Mirovský, J. *Proceedings of CIL17*. Praga : Matfyzpress, 2003. En: http://www.hf.uib.no/forskarskole/temmerman_art_prague03.pdf. Consultado el 14/12/2009.

TENOPIR, C. Trends in end user searching. *Library Journal*. Diciembre, nº 35-36. En: <http://www.libraryjournal.com/archive/>. Consultado el 15/12/2009.

TENOPIR, C. ; KING, D.W. Reading Behaviour and Electronic Journals. *Learned Publishing*, nº15, 2002.

Tesauros y glosarios iEDCYT en línea. En: www.cindoc.csic.es. Consultado el 15/12/2009.

TEZANOS, J. La crisis del franquismo y la transición democrática. En TEZANOS, J. ; COTARELO, R. ; BLAS, A. de.(eds.) *La transición democrática española*. Madrid: Sistema, 1989.

THE BRAIN. En: www.thebrain.com. Consultado el 15/02/2010.

The First International Topic Maps Users Conference. Oslo, 2007. En:

<http://www.topicmaps.com/tmc/conference.jsp?conf=TM2007>. Consultado el 15/02/2010.

THINKGRAPH. En: <http://www.thinkgraph.com>. Consultado el 15/12/2009.

THINKMAP VISUAL THESAURUS. En: <http://www.visualthesaurus.com>. Consultado el 15/12/2009.

THOMAS, PH. ; KNOWLES, J. *How to use a law library. An introduction to legal skills*. London : Sweet & Maxwell, 2001.

TILLET, B.B. AACR2 and Metadata: Library Opportunities in the Global Semantic Web. *Cataloging & Classification Quarterly*, vol.36, nº3/4, 2003. En: http://egovernment.cabinetoffice.gov.uk/Briefings/BriefingsArticle/fs/en?CONTENT_ID=4000207&chk=dtmcHu. Consultado el 25/02/2010.

TISCORNIA, D. *JurWordNet, semantic web of legal technology*. En: <http://www.ittig.cnr.it/Ricerca/UnitaEng.php?Id=11&T=4>. Consultado el 15/02/2010.

TISCORNIA, D. *The Lois Project: Lexical Ontologies for Legal Information Sharing*. En: <http://www.e-p-a-p.com/dlib/9788883980466/art14.pdf>. Consultado el 15/02/2010.

TMAPI. En: www.tmapi.org. Consultado el 15/12/2009.

TM4J. En: <http://tm4j.org/about.html>. Consultado el 15/10/2009.

TM4Jscript. En: <http://tm4jscript.sourceforge.net>. Consultado el 15/12/2009.

TMAPI.ORG. En: <http://tmapi.org>. Consultado el 15/12/2009.

TMCL. *Topic Map Constraint Language (ISO19756)*. En: <http://www.isotopicmaps.org/tmcl/>. Consultado el 15/07/2009.

TMPROC. En: <http://www.ontopia.net/software/tmproc>. Consultado el 15/10/2009.

TMTab. En: <http://www.techquila.com/tmtab.html>. Consultado el 15/12/2009.

TMWIKI. En: www.topic-maps.org. Consultado el 15/12/2009.

TOPIC MAPS. En:
http://www.wandora.org/wandora/wiki/index.php?title=Topic_Maps. Consultado el 15/02/2010.

TOPICMAPS.COM. En: www.topicmaps.com. Consultado el 15/12/2009.

TOPICMAPS.NET. En: www.topicmaps.com. Consultado el 15/12/2009

TOPICMAPS.ORG. En: www.topicmaps.org. Consultado el 15/12/2009.

TOPIC MAP THOUGHTS. En: <http://homepage.mac.com/dmitryv/iblog>. Consultado el 15/12/2009.

TRAMULLAS, J. ; GARRIDO PICAZO, P. Xml Topic Maps. Creating and using Topic Maps for the web. *El profesional de la información*, vol.14, nº 2, 2005.

TUDHOPE, D. ; ALANI, H. ; JONES, C. Aumenting Thesurus Relationships: Possibilities for Retrieval. *Journal of Digital Information*, nº 1, febrero 2001. En: <http://jodi.ecs.soton.ac.uk/Article/v01/i08/Tudhope>. Consultado el 27/12/09.

TVEIT, L. *How can local government use Topic Maps to make life easier for its citizens*. En: http://www.topicmaps.com/tmc/presentation.jsp?conf=TM2007&id=Lars_Tveit~TM2007~Bergen_Municipality. Consultado el 15/07/2009.

U

UNESCO. *Guidelines for the establishment and development of monolingual thesauri*. Paris, 1973.

UNIVERSIMMEDIA. En: <http://universimmedia.blogspot.com>. Consultado el 15/12/2009.

UREÑA LÓPEZ, L. *Resolución de la ambigüedad léxica en tareas de clasificación automática de Documentos*. Alicante: Sociedad Española para el Procesamiento del Lenguaje Natural, 2002.

V

VALENTE, A. Types and roles of legal ontologies. En: Benjamins, V.[et al.]. *Law and the Semantic Web. Legal ontologies, methodologies, legal information retrieval, and applications*. Berlin : Springer-Verlag, 2005.

VALENTE, A. ; BREUKER, J. Legal Modelling and Automated Reasoning with ON-LINE. *International Journal of Human-Computer Studies*, vol. 51nº6, 1999.

VATANT, B. Topic Maps from representation to identity. Conversation, names, and published subject indicators. En PARK, J. *XML Topic Maps. Creating and using topic maps for the web*. Boston : Addison-Wesley, 2002.

VLIST, Van Der. *Building a semantic web site*. En:
www.xml.com/pub/a/2001/05/02/semanticwebsite.htm. Consultado el 15/12/2009.

VAÑÓ VAÑÓ, Mª J. XML, una herramienta al servicio del buen gobierno corporativo. *Revista Aranzadi de Derecho y nuevas tecnologías*. Nº8, 2005.

VATANT, B. Binding points for Subject Identity: The case for standard Published Subject Indicators. En *Extreme Markup 2001*. En:
www.mulberrytech.com/Extreme/Proceedings/html/2001/Vatant01/EML2001Vatant01-toc.html. Consultado el 15/12/2009.

VATANT, B. *Building Knowledge Universe*. En:
http://www.mondeca.com/index.php/en/r_d/publications. Consultado el 15/12/2009.

VATANT, B. Managing complex environments with Topic maps. En *Knowledge Technologies Conference*, 2001. En:
<http://www2.gca.org/knowledgetechnologies/2001/proceedings/Vatant%20Paper.doc>
Consultado el 15/12/2009.

VATANT, B. Ontology-Driven Topic maps. En *Proceedings of XML Europe, 2004*. En: www.idealliance.org/europe/04/call/xmlpapers/03-03-03.91/.03-03-03.html. Consultado el 15/12/2009.

VATANT, B. OWL and Topic map Pudding. En *Mondeca White Paper, 2003*. En: www.mondeca.com/owl/owltm.html. Consultado el 14/12/2009.

VILLA BARAJAS, H. ; ALFONSO SÁNCHEZ, I. Biblioteca híbrida: el bibliotecario en medio del tránsito de lo tradicional a lo moderno. *Acimed*, vol. 13, nº2, 2005. En: http://bvs.sld.cu/revistas/aci/vol13_2_05/aci05_05.pdf. Consultado el 15/12/2009.

VISSER, P.R.S. *Knowledge specification for multiple legal tasks. A case study of the interaction problem in the legal domain*. The Hague : Kluwer Law International, 1995.

VOGEL, C. *Quality metrics: how to ensure quality taxonomies*. San Mateo: Semio Corporation, 1999.

VRANDECIC, D. ; PINTO, S. ; TEMPICH, C. ; SURE, Y. The DILIGENT knowledge processes. *Journal of Knowledge Management*, vol.9, nº5, 2005. En: http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2005_kmjjournal_diligent.pdf. Consultado el 15/02/2010.

W

WAHLSTER, W. (ed.). *Proceedings of the 12th European Conference on Artificial Intelligence (ECAI-96)*. Chichester : Wiley, 1996.

WARNER, A. Information architecture and vocabularies for browse and search. En GILCHRIST, A. ; MAHON, B. *Information Architecture. Designing information environments for purpose*. London : Facet, 2004.

WERSIG, G. *Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis*. München : Verlag Dokumentation Saur KG, 1978.

WIELINGA, B.J., SCHREIBER, A. T., BREUKER, J.A. KADS: A Modelling Approach to Knowledge Engineering. *Knowledge Acquisition*, vol. 4, n°1,1992.

WORDNET. En: www.cogsci.princeton.edu. Consultado el 16/12/2009.

WRIGHTSON, A. *Topic maps and Knowledge Representation*. En: <http://www.ontopia.net/topicmaps/materials/kr-tm.html>. Consultado el 15/02/2010.

WURMAN, R. *Information Architects*. New York : Graphics Press,1997.

WURMAN, S. *Angustia informativa*. Amsterdam : Prentice-Hall, 2001.

X

XML Conference & Exposition 2002. En: <http://62.231.133.220/xmlus02-stv/st.jsp?name=XML%202002>. Consultado el 15/12/2009.

Xsiteable. En: <http://xsiteable.sourceforge.net>. Consultado el 12/12/2009.

Y

YANG, H. ; LEE, Ch. Building Topic Maps using a text mining approach. En Zong, N. et al. (eds.). *ISMIS 2003*. Berlin : Springer, 2003.

YI, M. *Topic Maps-based Ontology and Semantic Web*. Saarbrücken : VDM Verlag, 2008.

YI, M. Information organization and retrieval using a topic maps-based ontology: Results of a task-based evaluation. *Journal of the American Society for Information Science and Technology*, vol. 59, n°12.

YU, S.-Ch. ; LU, K.-Y. ; CHEN, R.-S. Metadata management system: design and implementation. *The Electronic Library*, vol.21, n°2, 2003.

Z

ZANGERL, A. *TM View*. En: <http://search.cpan.org/~alphazulu/TM-View-0.5/>. Consultado el 12/01/2010.

ZEITLYN, D. ; BEX, J. Libraries without number. [*Special issue*]. *Education for Information*, vol.15, nº4, 1997.

ZELEZNIKOW, J. ; STRANIERI, A. An ontology for the construction of legal decision support systems. En *Proceedings of the second international workshop on legal ontologies*. Amsterdam, 2001. En: <http://www.lri.jur.uva.nl/jurix2001/papers/Zeleznikov.pdf>. Consultado el 12/12/2009.

ZTM *Topic maps*. En: <http://sourceforge.net/projects/ztm/>. Consultado el 12/12/2009.

APÉNDICE A.

GLOSARIO.

Associations

Representan cualquier tipo de relación entre los **topics**. Se trata de una representación de la relación entre una o más materias. Las *associations* cuentan con una *association type*, una materia que describe la naturaleza de la relación representada por las *associations* de ese tipo.

Un *association role* es una representación de la implicación de una materia en una relación representada por una *association*. Conecta dos fuentes de información dentro de una *association*: el *topic* que participa en la *association* y la materia que describe la naturaleza de la participación de ese mismo *topic* en la *association*.

CEPC.

Centro de Estudios Políticos y Constitucionales.

Clase

Un grupo, colección o conjunto de atributos que comparten alguna característica común. (Merriam Webster's Collegiate Dictionary).

Constraint o Gramática

Reglas que gobiernan las clases de objetos en una ontología con el fin de validar y/o crear de manera guiada un *topic map*. (ISO JTC1 SC34, N0226). La gramática es un poderoso mecanismo para el control en la creación de un *topic map* y la validación semántica del contenido del *topic map*. A esta gramática se la conoce por TMCL y la norma que la regula es ISO 19756: *Topic maps Constraint Language* (TMCL). Esta norma describe qué reglas pueden ser aplicadas para las distintas partes del *topic map* durante su creación.

Fuente de información

Representación de un recurso como secuencia de bytes susceptible de ser recuperada en una red digital como la web.

Identificador de materia

Localizador que se refiere a un indicador de materia.

Indicador de materia

Fuente de información referenciada desde un *topic map* con la intención de identificar sin ambigüedad la materia representada por un *topic* al usuario.

IRS Tax Map

Topic map de la agencia tributaria de Estados Unidos. Editada primitivamente para el *Call Centre* de la agencia de recaudación, terminó por ser una aplicación de asistencia a los ciudadanos.

Localizador de materia

Localizador que se refiere a aquella fuente de información que es materia de un *topic*.

Merging

Se trata de un proceso aplicado a los *topic maps* con el objeto de eliminar sus elementos redundantes. Esto supone que cuando dos *topic maps* tengan todos sus elementos en común, entonces deberán ser fusionados en uno.

Name

Es la denominación de un *topic*, consta de una forma base conocida por *base name* y las variantes de esa forma, conocidas por *variant names*. Un *topic name type* es una materia descriptiva de la naturaleza de los *topic names* de ese tipo. Los *names* siempre se definen en el ámbito o contexto más adecuado para la materia del *topic*, teniendo en cuenta que todo *topic* puede llegar a contar con varios *topic names*.

El *base name* se expresa como una cadena de caracteres que permite la identificación de una materia y que puede ser empleada como etiqueta en los interfaces de usuario. La noción de *base name* se corresponde de manera muy cercana con la denominada comúnmente el nombre.

NZETC

New Zealand Electronic Text Centre. Biblioteca digital de Nueva Zelanda editada con *Topic maps*.

Occurrences

Representan las fuentes de información relevantes para un **topic**. Se trata de una representación de la relación entre una materia y una fuente de información. La materia en cuestión es aquella representada por el topic que contiene la occurrence. La fuente puede ser bien un recurso interno al *topic map* como externo al mismo. En esencia las *occurrences* son un tipo de *association* especializada donde uno de los elementos implicados en la *association* es una fuente. Un *occurrence type* es una materia que describe la naturaleza de la relación entre las materias y las fuentes de información enlazadas por las *occurrences* de ese tipo.

Todas las *occurrences* tienen un **scope** o ámbito que define el contexto en el cual la relación *occurrence* entre la fuente y la materia es válida.

Ontología

Una ontología es una especificación explícita de una conceptualización. (T. Gruber, What is an Ontology?).

La ontología define los tipos de cosas que existen en el dominio de una aplicación. (J.F. Sowa, Knowledge Representation).

Reification

La acción de *reification* es el acto de hacer que un *topic* represente la materia de otro elemento de un *topic map* en el mismo *topic map*. Por ejemplo, la creación de un *topic* que represente la relación plasmada por una *association* es una *reification*.

Scope

Toda afirmación se realiza en un contexto. El *scope* representa ese contexto dentro del cual esa afirmación se considera válida. Fuera del contexto representado por el *scope*, la declaración esta no es válida. Formalmente, el *scope* se compone de una colección de *topics* que definen conjuntamente el contexto; es decir, la afirmación se considera válida únicamente en aquellos contextos donde se ven representadas todas las materias de la afirmación.

Subject

Un **subject** puede ser cualquier cosa, independientemente de su existencia o de sus características específicas, sobre la que pueda hacerse alguna declaración por cualquier medio.

Cualquier cosa que con independencia de su existencia o de sus características específicas, se pueda realizar una afirmación por cualquier medio posible.

***Template* o plantilla (topic, association, occurrence, scope)**

Declaración de una clase con sus reglas gramaticales respecto a las cuales sus instancias deben ser acordes. (ISO JTC1 SC34, N0226).

Topics

Representan cualquier concepto susceptible de ser representado por el hombre, desde conceptos abstractos hasta los más materiales. Es un símbolo que se utiliza dentro de un **topic map** para la representación de una materia mediante declaraciones que se hicieron sobre la misma. Una declaración es una afirmación acerca de una materia.

Topic Map

Es una norma para la representación e intercambio de conocimiento que incide en la localización de información. A esta norma ISO se la conoce como ISO/IEC 13250:2003.

Tecnología para la codificación de conocimiento y la conexión de este conocimiento codificado con sus fuentes de información pertinentes.

Colección de *topics* y *associations*.

Topic name

Denominación para un *topic* consistente de una forma base con sus correspondientes variantes denominadas *variant names*.

Topic map schema

Colección de plantillas que conjuntamente definen una clase de *topic map*.

Topic Type

Materia que representa alguna característica en común de entre un conjunto de materias.

XTM

XML para *Topic maps*. Lenguaje de marcado en XML definido para las características específicas de los *Topic maps*.